Psychometrika

CONTENTS

A FACTORIAL ANALYSIS OF THURSTONE'S SIXTEEN PRIMARY MENTAL ABILITIES - - - - - - 141
CHARLES H. GOODMAN

AN EXACT TEST OF SIGNIFICANCE FOR MEANS OF SAMPLES DRAWN FROM POPULATIONS WITH EXPONENTIAL FREQUENCY DISTRIBUTION - 153
LEON FESTINGER

ON THE RELIABILITY OF A WEIGHTED COMPOSITE - 161
CHARLES I. MOSIER

MULTIPLE RECTILINEAR PREDICTION AND THE RESOLUTION INTO COMPONENTS: II - - - - - - 169
LOUIS GUTTMAN AND JOZEF COHEN

A PREFERENTIAL MATCHING PROBLEM - - - - - 185
LT. J. A. GREENWOOD, U.S.N.R.

TABLES OF THE STANDARD ERROR OF TETRACHORIC CORRELATION COEFFICIENT - - - - - - 193
SAMUEL P. HAYES, JR.

VOLUME EIGHT SEPTEMBER 1943 NUMBER THREE

PSYCHOMETRIKA—VOL. 8, NO. 3
SEPTEMBER, 1943

A FACTORIAL ANALYSIS OF THURSTONE'S SIXTEEN PRIMARY MENTAL ABILITIES TESTS

CHARLES H. GOODMAN
THE PENNSYLVANIA STATE COLLEGE

This article deals with a factorial analysis of Thurstone's *Primary Mental Abilities Tests*. The analysis was made in order to determine whether the tests designated to measure a particular ability would be found upon completion of the analysis to be clustered together with significant loadings and be isolated from the tests of the other abilities. The results show that most of the tests specifically designed to measure six of the abilities were isolated with loadings varying in size. The tests for the remaining ability were not isolated. It is also shown that the tests are complex in nature, measuring more than any one single ability.

During the past twelve years, Professor L. L. Thurstone has conducted a considerable amount of research in the field of factor analysis. One of the many outcomes of these experiments has been the publication of a battery of tests designed to measure seven abilities. Thurstone (5) isolated these seven abilities as a result of his analysis of a battery of fifty-six tests which were given to a group of 250 University of Chicago students. These abilities were named: perception, number, verbal, space, memory, induction, and reasoning. The first five of these abilities are each measured by two tests, while the last two of these abilities are each measured by three tests, making a total of sixteen different tests in the battery. According to Thurstone (4), each test is heavily saturated with the particular ability that it is intended to measure.

During the fall of 1938 the Thurstone *Primary Mental Abilities Tests* (Experimental Edition, 1938) were administered to a group of 170 male freshmen engineering students at The Pennsylvania State College. Upon completion of the scoring process, the following question arose. Having presumably measured seven primary mental abilities, would it be possible to find the same seven abilities with which we started? Actually, the question is not one of verifying the composition of the seven primary abilities, but rather whether a factor analysis would isolate the sixteen primary mental abilities tests so that the particular tests designated to measure each ability would be found clustered together with significant projections for that factor. If we accept Thurstone's (4) statement that the tests are heavily

saturated with the abilities they were constructed to measure, it would seem, then, if the two or three tests designed to measure a particular ability were isolated and found to have significant loadings that these abilities do appear to be in the battery of tests. This study was planned to answer the foregoing question.

Subjects

The population used in the experiment consisted of 170 male freshmen taking the various engineering courses at The Pennsylvania State College. Their ages ranged from seventeen to twenty-one years, the mean age of the group being 19.58 years. These students were obtained as a result of an offer made to the entire freshmen engineering class of 278 students that those volunteering to take the tests would individually receive a psychograph of their abilities. It was thought that this procedure would result in greater cooperation. However, the question arose whether the students thus obtained would be a representative sampling of the entire freshmen engineering class. In order to determine whether the subjects of this study were representative of the freshman engineering class and the college population as a whole, a comprehensive investigation was made. The 170 subjects of this study served as the experimental group. The remaining 108 freshmen of the engineering class were used as a control group and the 778 freshmen enrolled in the various schools of the college served as a third group for the various comparisons that were made. The conclusions* of this ingestigation have been reported by the writer (3) in an earlier paper and show the subjects of this study to be a good representative sample of the freshmen engineering students and of the freshmen student body of The Pennsylvania State College.

Statistical Procedure

The subjects were assembled in groups of fifty and the Thurstone Primary Abilities Tests were administered. Due to the length of time, three hours and forty-one minutes of actual working time required to take the tests, it was decided to give the tests in two ses-

* A few aspects of this earlier study are of sufficient interest to note briefly. On the DeCamp (2) *Psychological Examination*, a test heavily weighted with general information items, and found to have a multiple R of .70 with college grade-point average, the mean DeCamp test score and sigma for the experimental group were not significantly different from the mean and sigma of the control group. Application of the Chi-Square test for goodness of fit showed the distribution of DeCamp test scores for the experimental group to be a good fit. The number of subjects found to be enrolled in each of the eight different engineering departments was closely similar in proportion to the numbers for the entire freshmen engineering class.

sions. Parts one and two were given in the first session and part three was given in the second session.

These abilities tests were constructed so that they could be scored by hand or electrical scoring machine. Scoring of the tests was done in this experiment by hand. First each test was completely scored; then as a check for accuracy each test was rescored by odd and even items. These latter data were used in computing the reliabilities for each of the tests and have been reported by Bernreuter and Goodman (1) in an earlier article.

The scores made by each of the subjects were punched into Hollerith cards and, with the aid of a tabulator, Pearson product moment coefficients were obtained between each of the sixteen tests.

These intercorrelations and their probable errors are shown in Table 1 and it was this table of correlations that served as the matrix table for the factor analysis. Thurstone's centroid method of factorial analysis was used. The data are analyzed to seven factors. The seventh factor residuals were examined for the largest positive and the largest negative entries in the new columns of the factorial matrix. None of these were found to be greater than +.10 and −.10. The maximum projection that could be obtained from rotation of the axes would then have been +.20. A projection of +.20 would account for only 4 per cent of the total variance of a test; hence there could be little assurance in the naming of a factor with so small a projection. Thus it was considered that a sufficient number of factors had been drawn. An attempt was made to apply Tucker's (5) empirical rule for determining the significant number of factors. However, the results were not clear enough to accept with assurance. The centroid analysis having been carried out, all of the factor loadings were collected into a single table, the factor matrix, and the communalities were computed for each of the tests. This is shown in Table 2. Then followed the process of rotation in order to make as many of the factor loadings as possible positive, and as many of the loadings as possible zero. The final outcome of this process of rotation is given in Table 3.

Results

Before examination of the outcome of the rotation of the axes, several items of interest can be noted in Table 1, the correlation matrix. It will be seen that the correlations between the tests of the factors range from −.11 to +.72. Table 4 is presented to show how the tests of each factor correlate with each other. The correlations between the tests of the number, verbal, and space factors are fairly

large, while the correlations of the perception, memory, induction, and reasoning factors are fairly small.

The tests of the factors perception, memory, induction, and reasoning correlate higher with tests which measure other factors than they do with tests that were designed to measure the same ability. Tests 1 and 2, which measure perception, correlate with each other .32. However, Test 1 has its highest correlation of +.48 with Test 7 of the space factor. Test 2 of perception has its highest correlation, +.43, with Test 6 of the verbal factor. Tests 9 and 10 of the memory factor correlate +.35, but Test 9 has its highest correlation, +.37, with Test 8 of the space factor, while Test 10 of memory also has its highest correlation, +.48, with Test 8 of the space factor. Of the three tests of induction, Test 11 correlates +.28 with Test 12 and +.33 with Test 13, while Test 12 correlates +.23 with Test 13. The highest correlation of Test 11 is +.38 with Test 15 of the reasoning factor. Test 12 has its highest correlation +.32 with Test 7 of the space factor. The highest correlation of Test 13 is +.33 with Test 11 of the same factor. The three tests of the reasoning factor correlate with each other in the following manner: Test 14 correlates +.44 with Test 15 and +.15 with Test 16. Test 15 correlates +.01 with Test 16. The highest correlation of Test 14, +.44, is with Test 15 of the same factor. Test 15 correlates highest, +.49, with Test 8 of the space factor, and Test 16 has its highest correlation, +.44, with Test 7 of the space factor. It is interesting to note that of the ten highest correlations for those tests measuring perception, memory, induction, and reasoning, six of them are with the two tests of the space factor. Of the 120 intercorrelations of the matrix table, 58 of these correlations appear to be significant when considered in terms of their P.E.'s. Nine of the 120 correlations are negative; however, they are small, ranging in size from −.01 to −.11.

The term "primary factor" has led to a great deal of confusion. Many have interpreted it to mean a pure psychological factor. In the light of such an interpretation questions would arise in connection with the intercorrelations of the matrix given in Table 1. Theoretically there should be no correlation between the tests of one factor and another if the primary abilities are pure. Wolfle (6) has clearly stated the problem in these words: "The factors defined by a simple structure are called *primary factors.* It should be emphasized that primary factors are defined mathematically; they do not necessarily have any genetic or psychological or philosophical primacy." Nevertheless, even from a factorial point of view the intercorrelations of the sixteen primary mental abilities tests would seem to indicate that

the individual tests are somewhat complex. Further evidence of this is found in the loadings of these tests for the various factors.

The final step of a factorial analysis is the psychological interpretation of the rotated factors. For this study the answer to our question of whether it is possible to isolate the tests of each factor lies in the rotated loadings of Table 3. Examination of the projections of Column I will quickly reveal that the two highest loadings are .80 and .76. Both of these projections are for the two tests which were designed to measure number ability. Following Thurstone's (5) procedure of not regarding a loading as significant unless it is as large as .40, it will be found that there are no other loadings in this column that are significant. Test 13, the Number Patterns test of the induction factor, has the next highest loading of .29 in the column. This would account for only 8 per cent of the total variance of the test. Although this loading is not significant, it is of interest to note that this test also deals with numbers. It will also be seen that the tests having the lowest projections in this column are preponderantly those involving verbal matter.

In Column II the three highest projections are .50, .48, and .41. All of these loadings belong to the three tests of the induction factor. Test 16, the Mechanical Movements test, designed to measure reasoning ability, has a projection of .39, which appears to be a significant loading. Thurstone (5) in his analysis of the battery of 56 tests which were given to the 240 Chicago University students, also found the Mechanical Movements test of that battery to have a significant projection of .40 for the deductive factor. There are two other projections of noticeable size in this column. The larger of the two is .38 for Test 7, the Cards test of the space factor. The smaller loading of .33 belongs to Test 14, the Arithmetic test of the reasoning ability. It is of interest to note that of the three loadings reported in this column other than those for the deductive factor, two of them are for tests of the reasoning ability.

The two largest loadings of Column III are for Tests 7 and 8, the Cards and Figures tests of the space factor. Both of these tests have loadings of .76. There are only two other loadings in this column worth noting: Test 10, the Initials test of the memory ability, has a loading of .37 and Test 1, the Identical Forms test of the perception factor, has a loading of .36. Examination of Thurstone's (5) rotated factor matrix for his battery of 56 tests will disclose that the Identical Forms test of that battery had a projection of .32 for the space factor. This is almost as large as the loading found in this study. It would appear from the task involved in the Identical Forms

test that a type of visualizing would be called into play. However, it does not seem to be the same sort of visual manipulation as that necessary for the solution of the tests measuring the space ability. A comparison of the loading found in this study for the Initials test of the Memory Ability with that of the loading found by Thurstone in his study shows a marked difference. The loading for the Initials test on the space factor in this study was .37 while Thurstone (5) reports a loading of −.09 for the Initials Test used in his battery. It is difficult to offer a satisfactory explanation why this test, which was designed to measure the ability to memorize, would appear on the space factor, which calls for the ability to be able to think visually of geometrical forms in space.

In Column IV the interpretation of the loadings appears to be very clear. There are only two loadings that can be regarded as significant. Both of these projections are for Tests 9 and 10, the Word Number test and Initials test of the memory ability.

Tests 5 and 6, the Completion test and Same Opposites test, which both measure the verbal ability, have the highest projections of .68 and .71, respectively, in Column V. There are two other loadings in this column that should be noted. The first is Test 2, the Verbal Enumeration test of the perception ability, which has a projection of .43, which is, according to Thurstone's criterion, significant. Examination of Test 2 will disclose that despite the fact that it presumably measures perception, it is heavily loaded with verbal material and it would appear quite plausible that it should have a significant loading on the verbal factor. The other test having a noticeable loading in this column is Test 3, the Additions test of the number ability. It is difficult to comprehend what the underlying relationship might be in this instance.

The loadings of Column VI present a rather clear picture. There are only two projections that can be considered significant. Both of these loadings are for tests which measure the reasoning ability. Test 14, the Arithmetic test, has a projection of .51, while Test 15, the Number Series test, has a projection of .58. Thurstone devised three tests to measure this reasoning ability, but it can be seen from the projections of this column that only two of the tests have shown up significantly. The third test measuring this ability is Mechanical Movements, and it has a loading of −.06 in this column. According to the data presented here, it would appear that the Mechanical Movements test is not measuring the same ability as that measured by the tests of Number Series and Arithmetic. It will be recalled that the Mechanical Movements test correlated only +.01 with the Number

Series test and +.15 with the Arithmetic test.

There are no loadings of any significance in Column VII. The highest loading to be found is .32 for Test 8, the Figures test of the space factor. It would be difficult to attempt to name the factor that might exist in this column with loadings of this size.

It may be noted that this analysis did not identify Thurstone's perceptual factor, which he has withdrawn from the *Primary Mental Abilities Tests* pending further study of this factor.

Summary

This study of the factorial analysis of Thurstone's sixteen primary mental abilities tests based on the first experimental edition, would seem to justify the following conclusions:

(1) The factor analysis disclosed that it was possible to isolate the two or three tests which measured each factor except perception.

(2) The particular tests designated to measure each of the six abilities of number, induction, space, memory, verbal, and reasoning, were found to be clearly isolated, clustered together, with significant factor loadings.

(3) The Mechanical Movements test, one of the three tests designated to measure reasoning ability, failed to have a significant loading for that ability.

(4) The loadings of the tests isolated for each of the abilities vary in size from .41 to .80. This would account for a range of from 16 to 64 per cent of the total variance of the tests.

(5) The correlation of some of the tests are higher with tests which measure other factors than they are with the tests that were designed to measure the same ability.

(6) The correlations between the two or three tests which measure each of the particular abilities range between .01 and .72.

(7) From a factorial viewpoint, some of the *Primary Mental Abilities Tests* can be regarded as being somewhat complex. This can be seen from the loadings these tests have on the various factors which indicates that more than one factor is involved in the test.

TABLE 1

Table Showing the Intercorrelations of the Sixteen Primary Mental Abilities Tests

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	+.32	+.16	+.28	+.21	+.17	+.48	+.29	+.03	−.02	+.19	+.20	+.18	−.01	+.12	+.17
	2	+.17	+.27	+.34	+.43	+.22	+.23	+.09	±.00	+.08	+.16	+.20	−.01	−.03	+.18
		3	+.72	+.22	+.28	+.24	+.24	+.04	+.05	+.02	+.06	+.27	+.24	+.25	−.11
			4	+.21	+.26	+.24	+.34	+.16	+.03	+.19	+.11	+.25	+.20	+.19	−.05
				5	+.61	+.20	+.20	+.10	+.11	+.16	+.15	+.23	+.31	+.31	+.18
					6	+.18	+.20	+.17	+.12	+.08	+.10	+.13	+.25	+.20	+.18
						7	+.63	+.11	+.20	+.26	+.32	+.27	+.15	+.17	+.44
							8	+.37	+.48	+.17	+.30	+.23	+.10	+.49	+.23
								9	+.35	+.10	+.04	+.12	+.08	+.10	−.08
									10	+.14	+.04	+.11	+.01	+.06	−.03
										11	+.28	+.33	+.34	+.38	+.08
											12	+.22	+.22	+.17	+.26
												13	+.20	+.21	+.13
													14	+.44	+.15
														15	+.01
															16

r	P.E.
.60-.79	.024
.40-.59	.035
.20-.39	.046
.00-.19	.052

TABLE 2
Original Factor Loadings and Communalities

Number	Tests	Original Factor Loadings							Communalities
		I	II	III	IV	V	VI	VII	h^2
1	Identical Forms	+.4433	.1342	.0502	−.3235	.2184	−.0583	−.1549	.3967
2	Verbal Enumeration	.4184	−.0765	−.0291	−.1949	.3455	−.2519	.2071	.4454
3	Addition	.4884	−.5083	.4864	−.1715	−.1546	.2037	−.0753	.8340
4	Multiplication	.5639	−.3555	.4553	−.2359	−.1484	−.0510	.1340	.7499
5	Completion	.5629	−.3349	−.2762	.1924	.3015	−.0560	−.1210	.6510
6	Same Opposites	.5422	−.3434	−.1320	.1850	.4014	−.0865	.1133	.6449
7	Cards	.6437	.4491	.0881	−.2601	.1550	.2449	−.1220	.7903
8	Figures	.6945	.3952	.2891	.2524	−.0409	.1950	.0579	.8287
9	Word Number	.2959	.1372	.1949	.3500	−.0546	−.2614	.1599	.3638
10	Initials	.2952	.3029	.2214	.4044	−.0717	−.2240	−.0613	.4504
11	Letter Grouping	.4335	.0554	−.2388	−.0475	−.3426	−.1600	−.1798	.4256
12	Marks	.4060	.1907	−.2294	−.1713	−.1339	.0656	.1181	.3192
13	Number Patterns	.4673	−.0091	−.0800	−.1673	−.2112	−.1382	−.1346	.3347
14	Arithmetic	.4260	−.2176	−.3061	.1321	−.3078	.2125	.0419	.4816
15	Number Series	.4847	−.0819	−.1130	.2770	−.3063	.2530	−.1607	.5147
16	Mechanical Movements	.2973	.2293	−.3006	−.1918	.2397	.2541	.1652	.4176

TABLE 3
Rotated Factor Loadings and Communalities

Number	Tests	Rotated Factor Loadings							Communalities
		I^{V}	II^{IV}	III^{V}	IV^{III}	V^{II}	VI^{VI}	VII^{I}	h^2
1	Identical Forms	.1221	.2997	.3631	.0331	.2718	−.2415	−.1637	.3965
2	Verbal Enumeration	.0531	.2910	.0594	.2150	.4393	−.3109	.1355	.4452
3	Addition	.8029	−.1187	.1555	−.0842	.3706	.0764	.0268	.8338
4	Multiplication	.7653	.0772	.1310	.1224	.2963	−.1166	.1574	.7500
5	Completion	−.0154	.2193	.0605	.2163	.6872	.2808	−.0247	.6504
6	Same Opposites	.0091	.1348	.0398	.2539	.7162	.1052	.1897	.6444
7	Cards	.0953	.3850	.7640	−.0981	.1609	.1183	.0052	.7905
8	Figures	.1893	.1285	.7659	.2262	.0494	.1796	.3225	.8287
9	Word Number	.0737	−.0119	.1892	.5021	−.0262	.0744	.2523	.3633
10	Initials	.0014	−.0488	.3742	.5155	−.1191	.1439	.0845	.4502
11	Letter Grouping	.2025	.4846	.0997	.1922	−.0269	.2710	−.1661	.4243
12	Marks	.0686	.5026	.1986	−.0454	.0159	.0618	.1280	.3192
13	Number Patterns	.2985	.4102	.1428	.1462	.0873	.0966	−.1361	.3346
14	Arithmetic	.2062	.3377	−.0077	−.0637	.1655	.5194	.1520	.4810
15	Number Series	.2287	.1766	.2524	.0297	.1416	.5866	.0497	.5147
16	Mechanical Movements	.2179	.3968	.2589	−.2464	.2119	−.0677	.1881	.4175

TABLE 4

Table Showing the Correlations Between the Tests Which Measure Each of the Seven Factors

Factor	No.	Test Measuring Factor	Coefficient of Correlation	P.E.
P	1	Identical forms	+.32	.05
	2	Verbal Enumeration		
N	3	Addition	+.72	.02
	4	Multiplication		
V	5	Completion	+.61	.03
	6	Same Opposites		
S	7	Cards	+.63	.03
	8	Figures		
M	9	Word Number	+.35	.05
	10	Initials		
I	11	Letter Grouping	+.28 (11 & 12)	.05
	12	Marks	+.33 (11 & 13)	.05
	13	Number Patterns	+.22 (12 & 13)	.05
D	14	Arithmetic	+.44 (14 & 15)	.04
	15	Number Series	+.15 (14 & 16)	.05
	16	Mechanical Movements	+.01 (15 & 16)	.05

REFERENCES

1. Bernreuter, Robert G., and Goodman, Charles H. A study of the Thurstone Primary Mental Abilities Tests applied to freshman engineering students. *J. educ. Psychol.*, 1941, **32**, 55-60.
2. DeCamp, J. E. *The Pennsylvania State College Psychological Examination*, The Pennsylvania State College Press, State College, 1938.
3. Goodman, Charles H. *A study of the Thurstone Primary Mental Abilities Tests.* State College, Pennsylvania: The Pennsylvania State College, 1941. (Unpublished).
4. Thurstone, L. L. *Manual of instructions for the Primary Mental Abilities Tests.* American Council on Education, Washington, D. C., 1938.
5. Thurstone, L. L. Primary mental abilities. *Psychometric Monographs*, No. I, Univ. Chicago Press, 1939.
6. Wolfle, Dael. Factor analysis to 1940. *Psychometric Monographs*, No. 3, Univ. Chicago Press, 1940.

PSYCHOMETRIKA—VOL. 8, NO. 3
SEPTEMBER, 1943

AN EXACT TEST OF SIGNIFICANCE FOR MEANS OF SAMPLES DRAWN FROM POPULATIONS WITH AN EXPONENTIAL FREQUENCY DISTRIBUTION

LEON FESTINGER

STATE UNIVERSITY OF IOWA*

The mathematical derivation of a test for determining the fiducial limits of, and significance of difference between, means when the samples are drawn from exponential populations is presented. The test for differences between means takes the particularly simple form of the F test (the ratio of the larger to the smaller mean) with each mean possessed of $2n$ degrees of freedom, n being the number of cases in the sample. Random sampling, a range of scores upwards from a lower limit of zero, and independence of means from each other are necessary assumptions for the use of the test. Examples of situations in which the test should be used are given, together with a description of the necessary computational procedures. Comparisons of the results of the application of this test with the erroneous application of the critical ratio on actual data show that rather large discrepancies exist between the two tests. Results obtained by applying tests which assume normality to exponential distributions are subject to much error.

The research psychologist is frequently confronted with data where the frequency distributions are exponential in character. Such distributions occur, for example, in studies involving cultural or institutional behavior, where there is a maximum frequency at the cultural norm and a tapering off of frequency in one direction away from this norm (J-curves) and in time observation studies of individuals and groups where for many behavior items there is a maximal frequency at zero occurrence within a given time interval and the frequency decreases progressively as one moves out to greater and greater numbers of occurrences per unit time.

As yet, when dealing with such distributions there has been no method for testing hypotheses as to the value of the true mean or for testing significance of differences between means. The t test is obviously incorrect in such instances, since even for large samples the means do not distribute themselves normally. The present paper presents the mathematical derivation, and some applications, of an exact test for means of samples from exponential populations.

* From the Iowa Child Welfare Research Station.

I. Fiducial limits of the population mean:

Let us assume a population with an exponential frequency function, the general form of which may be written:

$$f(x) = ce^{-bx}; \qquad 0 \leq x < \infty . \tag{1}$$

The constants b and c of this function may be expressed in terms of moments of the distribution. Thus, setting

$$\int_0^\infty f(x)\,dx = k,$$

and

$$\frac{1}{k}\int_0^\infty xf(x)\,dx = x_p,$$

and solving, we find $c = kb$ and $b = 1/x_p$, where x_p is the population mean and k is the total area under the curve. We may then write equation (1) as:

$$f(x) = \frac{k}{x_p}e^{-x/x_p}; \qquad 0 \leq x < \infty. \tag{1a}$$

Craig* (2) has demonstrated that for a given population represented by the distribution function (1a), the arithmetic means (x_s) of samples of size n will distribute themselves in accordance with:

$$\phi(x_s) = \frac{(kn)^n\, x_s^{n-1}}{x_p^{n}(n-1)!}e^{-nx_s/x_p}. \tag{2}$$

To get the probability distribution of x_s we have merely to set $k = 1$. By regrouping of terms and setting $k = 1$, we may write equation (2) in a more convenient form:

$$\phi(x_s) = \frac{\left(\frac{nx_s}{x_p}\right)^{n-1} e^{-(nx_s/x_p)}}{\frac{x_p}{n}(n-1)!}. \tag{2a}$$

Equation (2a), then, represents the distribution of x_s.

Let us find the distribution function of z, where $z = \dfrac{2nx_s}{x_p}$. We may then write:

* Irwin (4) has, by a different method, found the same distribution function for means from an exponential population, thus checking the results of Craig. In Irwin's derivation the distribution function of means from an exponential population comes out as a special case of the distribution of means from a Pearson Type III population.

$$dz = \frac{2n}{x_p}\, dx_s, \tag{3}$$

and

$$dx_s = \frac{x_p}{2n}\, dz. \tag{3a}$$

The integral of $f(x_s)$ from x_{s1} to x_{s2} must equal the integral of $f(z)$ from $\frac{x_p}{2n} z_1$ to $\frac{x_p}{2n} z_2$. We may then write:

$$\frac{1}{\frac{x_p}{n}(n-1)!} \int_{x_{s1}}^{x_{s2}} \left(\frac{nx_s}{x_p}\right)^{n-1} e^{-(nx_s/x_p)}\, dx_s$$

$$= \frac{\frac{x_p}{2n}}{\frac{x_p}{n}(n-1)!} \int_{x_p/2n\, z_1}^{x_p/2n\, z_2} \left(\frac{z}{2}\right)^{n-1} e^{-(z/2)}\, dz; \tag{4}$$

then:

$$F(z) = \frac{\left(\frac{z}{2}\right)^{n-1} e^{-(z/2)}}{2(n-1)!} = F\left(\frac{2nx_s}{x_p}\right). \tag{5}$$

Let us now turn our attention for a moment to the distribution function of Chi Square. This function, as given by Kenney (5), is:

$$F(x^2) = \frac{\left(\frac{x^2}{2}\right)^{k-1/2-1} e^{-(x^2/2)}}{2\,\Gamma\left(\frac{k-1}{2}\right)} \quad \text{for } (k-1) \text{ degrees of freedom.} \tag{6}$$

It may be observed that equations (5) and (6) are identical [n in equation (5) equals $\frac{1}{2}(k-1)$ of equation (6)]. In other words, $2x_s/x_p$ distributes itself as Chi Square for $2n$ degrees of freedom.

It is then obviously possible to test any hypothesis as to the value of the true mean, given a sample mean from an exponential population, by referring $2nx_s/x_p$ to the table of Chi Square for $2n$ degrees of freedom. For n greater than 15, $\sqrt{4nx_s/x_p}$ distributes itself approximately normally about $\sqrt{4n-1}$ with unit variance and so the quantity $\sqrt{4nx_s/x_p} - \sqrt{4n-1}$ may be referred to the normal probability scale.

Let us work through an illustration of each of the above cases (n less than 15, and n greater than 15) to show more clearly the usage of this test.

Example for $n < 15$

A sample of 6 cases drawn at random from an exponential population. The mean of the sample is 10. The problem is to establish the 1 per cent and 5 per cent level fiducial limits of the true mean. We find that Chi Square, for 12 ($2n$) degrees of freedom, must equal 3.571 at the lower limit and 26.217 at the upper limit for significance at the 1 per cent level of confidence. Then $2nx_s/x_p$ must equal, in turn, 3.571 and 26.217. Then $120/x_p = 3.571$ gives the upper fiducial limit of 33.60; $120/x_p = 26.217$ gives the lower fiducial limit of 4.58. At the 1 per cent level of confidence, 4.58 and 33.60 are the limits between which we may legitimately entertain any hypothesis about the value of the true mean.

Likewise, the 5 per cent levels of Chi Square at the two ends of the distribution are 5.226 and 21.026. Setting $120/x_p$ equal to each of these in turn, we obtain the limits of 5.71 and 22.96 within which, at the 5 per cent level of confidence, we may entertain hypotheses about the value of the true mean.

Example for $n > 15$

A sample of 20 cases drawn at random from an exponential population. The mean of the sample is 10. The problem is the same as in Example 1. The $\sqrt{4n - 1}$ in this case equals 8.888. $\sqrt{4nx_s/x_p}$ in this case equals $28.284/\sqrt{x_p}$. The normal probability table gives 1 per cent of the cases lying outside the mean plus or minus 2.58. We may then set $28.284/\sqrt{x_p - 8.888} = \pm 2.58$. Solving for x_p we obtain fiducial limits of 6.08 and 20.11 within which we may entertain hypotheses as to the value of the true mean at the 1 per cent level of confidence.

Similarly, using the value 1.96 instead of 2.58 in the equations above, and solving, we obtain the 5 per cent fiducial limits of 6.80 and 16.67.

It is important to note that the fiducial limits are not symmetrical about the obtained mean.

It should be clear that for the test to hold, the population must vary from zero to infinity. If samples are obtained which are exponential in character, but where the lowest possible score is some constant $c > 0$, then the matrix must be transposed so that $c = 0$. For the purposes of the test we have merely to take $(x_s - c)$.

II. Significance of difference between means:

It is important at this point to consider the distribution of the variance of samples of size n drawn from a normal population. This distribution, as given by Kenney (5), is:

$$H(s^2) = \frac{e^{-(Ns^2/2\sigma^2)}\left(\frac{Ns^2}{2\sigma^2}\right)^{(N-3)/2}}{\left(\frac{2\sigma^2}{N}\right)\Gamma\left(\frac{N-1}{2}\right)}, \tag{7}$$

where N is the number of cases in the sample, s^2 is the variance of the sample, and σ^2 is the variance of the population. It may be observed that there is a similarity between equation (7) and equation (2). In fact, the quantity $\frac{Ns^2}{\sigma^2} = \frac{\hat{n\sigma^2}}{\sigma^2}$ (where $n = N - 1 =$ the degrees of freedom and $\hat{\sigma^2}$ is the estimate of the population variance from the sample) is distributed as Chi Square for n d.f.. In other words:

$$F\left(\frac{n\hat{\sigma^2}}{\sigma^2}\right) = F\left(\frac{2nx_s}{x_p}\right) = F(X^2). \tag{8}$$

Fisher's z-distribution, then, holds for the logarithm of the ratio of means drawn from the same exponential population for degrees of freedom equal to $2n_1$ and $2n_2$ just as it holds for the logarithm of the ratio of variance estimates drawn from a normal population for degrees of freedom equal to $N_1 - 1$, and $N_2 - 1$. In other words, the ratio of the means from an exponential population may be referred to the table of F values, for the appropriate number of degrees of freedom, for a test of significance between them. Thus:

$$\frac{x_{si}}{x_{sj}} = F; \quad \text{degrees of freedom} = 2n_i \text{ and } 2n_j. \tag{9}$$

III. Applications:

The following are some cases, drawn from psychological literature, where exponential distributions have been found. They will be discussed below to illustrate the applications of the test proposed in this paper.

Case 1

Dudycha (3) made observations on college students' coming early or late in various situations. One of the measures computed for each person in each situation is the ratio of "number of times early"

multiplied by 100 divided by the "number of times present." The frequency distributions of this measure are presented [(3), Fig. 7, p. 25] and are obviously exponential in character. The individual scores range from 100 per cent early to zero per cent early with the maximum frequency at the midpoint of the interval 100 - 95, and a tapering off from there on.

Since our test requires the distributions to taper off from a maximum at zero, we must, for the purposes of the test, transpose the scores by means of (100 − score) so that zero per cent is called 100 per cent, and vice versa.

Selecting two of the distributions, namely, promptness at breakfast ($N = 110$) and promptness at 8:00 o'clock classes ($N = 211$), we find the means to be 5.4 per cent and 27.6 per cent, respectively. The author presents no test of significance between these two distributions, but we may ask ourselves whether these two means do or do not differ significantly. Since the F necessary at the 1 per cent level is 1.33 for these magnitudes of degrees of freedom, it is apparent that the ratio of the larger to the smaller mean ($27.6/5.4 = 5.11$) is significant beyond the 1 per cent level of confidence. We may then conclude that the two means are significantly different.

Case 2

Schanck (6) asked 61 people in a community to rate their attitude toward prohibition on a 4-point continuum ranging from "favor enforcement" to "favor modification." The ratings were done in public and again in private. The distributions are of the type known as J-curves with the maximum frequency at "favor enforcement."

The four points on the scale may be given numerical values. We shall regard the first point as the midpoint of the interval from zero to 1 and assign to it the value of 0.5, to the next point the value of 1.5, and so on. When this is done, the mean of the distribution for ratings in public is found to be .975, and the mean for ratings in private is found to be 1.189.

The fiducial limits of the true mean may be found for either of these distributions. It is important to note, however, that the F test cannot be applied to test the significance of difference between the means since they both represent data on the same people and are therefore not independent. A test could be applied to the distribution of differences between the public and private ratings.

Case 3

Arrington (1) observed how frequently children spoke to other children in 5-minute periods of observation. The observations were

made in the normal preschool and kindergarten situation. The frequency distributions [(1), Figs. 4 and 5, pp. 74, 75] are obviously exponential in character. The means obtained and the number of cases in each sample are given in the tabulation below.

	Boys		*Girls*	
	N	Mean	*N*	Mean
Younger nursery	19	2.61	17	2.22
Older nursery	9	6.96	11	7.59
Kindergarten	26	12.76	24	10.06

In order to determine whether the mean number of times a child speaks to another child increases significantly as the children grow older, the author calculated the standard deviations of the distributions and tested significance of difference between means with the critical ratio. For the difference between younger and older nursery school children, for boys and girls, respectively, the critical ratios obtained are 8.37 and 10.74. When the correct procedure of applying the *F* test to the ratio of the larger to the smaller mean is used, the *F*'s are 2.667 and 3.419, respectively. The difference is significant at the 1 per cent level since the values of *F* required for such significance, for the degrees of freedom involved, are 2.45 and 2.42, respectively.

The differences between the older preschool children and the kindergarten children yield critical ratios of 9.20 and 6.53, respectively, for boys and girls. The correct *F* test yields *F* values of 1.833 for boys and 1.325 for girls, neither of which is significant even at the 5 per cent level. The values of *F* that would be necessary at the 5 per cent level, for the degrees of freedom involved, would be 2.04 and 1.91, respectively. We must then conclude, contrary to the critical ratios, that these latter differences could very well have been obtained by chance.

The present test makes available a correct method of analysis for means, and differences between means, of samples drawn at random from exponential populations. The test is easy to apply and should prove of value to the psychologist.

To summarize, we shall list below the assumptions on which the tests are based:

(1) The parent population has an exponential frequency distribution.
(2) The sample is drawn at random.
(3) The scores of the parent population range upward from zero.

(4) If we are testing the significance of difference between means, the means are assumed to be statistically independent.

REFERENCES

1. Arrington, R. E. Time-sampling studies of child behavior. *Psychol. Monog.*, 1939, 51, No. 228, 3-193.
2. Craig, A. T. On the distributions of certain statistics. *Amer. J. Math.*, 1932, 54, 353-366.
3. Dudycha, G. J. An objective study of punctuality in relation to personality and achievement. *Arch. Psychol.*, 1936, 29, No. 204, 1-53.
4. Irwin, J. O. On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's Type II. *Biometrika*, 1927, 19, 225-239.
5. Kenney, J. F. Mathematics of statistics, Part Two. 1939. New York: D. Van Nostrand Co. Pp. 202.
6. Schanck, R. L. A study of a community and its groups and institutions conceived of as behavior of individuals. *Psychol. Monog.*, 1932, 43, No. 195, 1-133.

PSYCHOMETRIKA—VOL. 8, NO. 3
SEPTEMBER, 1943

ON THE RELIABILITY OF A WEIGHTED COMPOSITE*

CHARLES I. MOSIER
SOCIAL SECURITY BOARD

A general formula for the reliability of a weighted composite has been derived by which that reliability can be estimated from a knowledge of the weights whatever their source, reliabilities, dispersions, and intercorrelations of the components. The Spearman-Brown formula has been shown to be a special case of the general statement. The effect of the internal consistency or intercorrelation of the components has been investigated and the conditions defining the set of weights yielding maximum reliability shown to be that the weight of a component is proportional to the sum of its intercorrelations with the remaining components and inversely proportional to its error variance.

Although the usual purpose of weighting the component parts of a test battery in combining them to form a composite total score is to increase the validity of the composite, there may be times when it is desired to assign weights in such a way as to produce the most reliable composite. In the absence of a measurable external criterion, for instance, the maximum reliability of the composite is certainly as legitimate a basis for assigning weights as is the intuitive assignment of small whole numbers. In any event, even if the weights are assigned in such a way as to provide a least-squares estimate of some quantitative criterion, the investigation of the effect of the weighting of the components on the reliability of the composite is a problem of considerable, even if secondary, importance.

The present paper develops a generalized formula for the reliability of a weighted composite, investigates certain special cases of unusual interest, and derives the conditions of maximum reliability.†

The Generalized Formula for the Reliability of a Composite

The generalized statement of the reliability of a composite is a function of the reliabilities, the dispersions, and the intercorrelations of the components. It is here derived and stated in both explicit and in parametric form. Let us suppose a composite variable y, consisting

* The opinions expressed in this article are those of the author and do not necessarily reflect the official views of the Social Security Board.

† Certain of the deductions set forth in this paper are implicit in Kelley, T. L., *Statistical Method*, Chapter IX, and others in Richardson, M. W. "The Combination of Measures," in Horst, Paul, *The Prediction of Personal Adjustment*, Social Science Research Council, 1941.

of the weighted sum of n component scores, $x_1, x_2, \cdots, x_j, \cdots, x_n$, so that for any individual, i,

$$y_i = W_1 x_{i1} + W_2 x_{i2} + \cdots + W_j x_{ij} + \cdots + W_n x_{in}, \tag{1}$$

where both y_i and x_i are deviation scores.

Let us represent the variance due to error of the j-th component by its squared standard error, given by

$$\varepsilon^2_j = \sigma^2_j (1 - r_{jj}), \tag{2}$$

where r_{jj} is the reliability of the component x_j. Now, since the error variance of each x_j is multiplied, in its contribution to the error variance of the composite, by W_j and since by definition the error variances of the components are uncorrelated, we may write the error variance of y as

$$\varepsilon^2_y = \sum_j^n W^2_j \varepsilon^2_j = \sum_j^n W^2_j \sigma^2_j (1 - r_{jj}) = \sum_j^n W^2_j \sigma^2_j - \sum_j^n W^2_j \sigma^2_j r_{jj}. \tag{3}$$

Now the total variance of the composite y is given by the expression

$$\begin{aligned} \sigma^2_y &= W^2_1 \sigma^2_1 + W^2_2 \sigma_2{}^2 + \cdots + W^2_j \sigma^2_j + \cdots + W^2_n \sigma^2_n \\ &\quad + \cdots + 2 r_{1j} W_1 W_j \sigma_1 \sigma_j + \cdots + 2 r_{jk} W_j W_k \sigma_j \sigma_k \qquad (4) \\ &= \sum_j^n W^2_j \sigma^2_j + 2 \sum_j^n \sum_k^n W_j W_k \sigma_j \sigma_k r_{jk}. \qquad (j < k) \end{aligned}$$

But the reliability of y may be expressed as a function of the ratio of the error variance to the total variance, thus

$$\begin{aligned} r_{yy} &= 1 - \frac{\varepsilon^2_y}{\sigma^2_y} \\ &= 1 - \frac{\sum W^2_j \sigma^2_j - \sum W^2_j \sigma^2_j r_{jj}}{\sum_j^n W^2_j \sigma^2_j + 2 \sum_{j=1}^{n-1} \sum_{k=2}^{n} W_j W_k \sigma_j \sigma_k r_{jk}}. \qquad (j < k) \end{aligned} \tag{5}$$

This, then, is the basic formula by which the reliability of the weighted composite may be estimated from a knowledge of the dispersions, the weights, the reliabilities, and the intercorrelations of the components. The W_j may be the weights obtained from a least-squares regression equation, or the small whole numbers assigned on some arbitrary or intuitive basis to increase the "validity" with which the composite will predict some vaguely defined, unmeasured "criterion." In

either case the reliability of the resulting composite may be estimated from equation (5).

It is convenient, and in accord with common practice, to simplify equation (5) without loss of generality by the reduction of the x_j to standard scores and the introduction of the parametric equation.

$$w_j = W_j \sigma_j \tag{6}$$

so that (5) becomes

$$r_{ww} = 1 - \frac{\sum_{j=1}^{n} w^2_j - \sum_{j=1}^{n} w^2_j r_{jj}}{\sum_{j=1}^{n} w^2_j + 2\sum_{j=1}^{n-1}\sum_{k=2}^{n} w_j w_k r_{jk}}, \tag{7}$$

remembering that $j < k$. Equation (7) may also be written as (8)

$$r_{ww} = \frac{\sum_{j}^{n} w^2_j r_{jj} + 2\sum_{j=1}^{n-1}\sum_{k=2}^{n} w_j w_k r_{jk}}{\sum w^2_j + 2\sum_{j=1}^{n-1}\sum_{k=2}^{n} w_j w_k r_{jk}}. \tag{8}$$

Equation (8) is of especial interest, since it can readily be reduced to matric form. Let us define a row vector, W_j, on the range j, consisting of the n weights, w_j; let the matrix of intercorrelations of the components be r, where r_{jj} is, as defined above, the reliability coefficient of the test j; let us also define a second matrix R so that each element $R_{jk} = r_{jk}$, but the diagonal elements, $R_{jj} = 1 \neq r_{jj}$. Then the numerator of (8) is seen to be $W_j\, r\, W'_j$ and the denominator is seen similarly to be WRW', the expressions for numerator and denominator differing only in the diagonal elements of the matrix of intercorrelations, since these are reliability coefficients in the numerator term and total variances in the denominator term. Equation (8) written in matric notation is thus seen to be:

$$r_{ww} = W\, r\, W' (W\, R\, W')^{-1}. \tag{9}$$

Special Cases and General Interpretations. Before determining the values of w_j which make r_{ww} a maximum, it may be of interest to investigate certain special cases and general properties of the formulation. The most obvious conclusion is, of course, that r_{ww} is unity if, and only if, *every* r_{jj} for which w_j is not zero is also unity—a conclusion of some consequence for the practice of attempting to increase reliability solely by adding items. Similarly, if every r_{jj} is zero, with its

corollary that every r_{jk} is zero, then it follows that r_{yy} is zero—not a very profound conclusion.

Concerning the values of the variables, w_j, r_{jk} and r_{jj}, we may make certain assumptions and investigate the effect of these special conditions. This has been done systematically for equal and unequal weights, equal and unequal reliability coefficients, and for r_{jk} equal to zero, unity, and unity when corrected for attenuation, i.e., $r_{jk} = \sqrt{r_{jj} r_{kk}}$. Only those cases of particular interest are reported here.

The Spearman-Brown Formula. A special case of particular interest is the Spearman-Brown formula. If not only the dispersions are equal [or equated as in (6)], but the weights of the components are equal, and set for convenience equal to unity, and if, in addition, the reliabilities are also equal, equation (7) becomes

$$r_{yy} = 1 - \frac{n - n r_{jj}}{n + 2 \sum_{j=1}^{n-1} \sum_{k=2}^{n} r_{jk}}. \tag{10}$$

But remembering that $j < k$, and letting $\bar{r}_{jk}$ be the average value of r_{jk},

$$2 \sum_{j=1}^{n-1} \sum_{k=2}^{n} r_{jk} = n(n-1)\bar{r}_{jk}. \tag{11}$$

and making this substitution in equation (10), we have

$$r_{yy} = 1 - \frac{n(1 - r_{jj})}{n + n(n-1)\bar{r}_{jk}} = 1 - \frac{1 - r_{jj}}{1 + (n-1)\bar{r}_{jk}}. \tag{12}$$

Equation (12) not only constitutes a lemma convenient in the derivation to follow, but contains certain conclusions of importance, to which reference will shortly be made.

If the intercorrelations of the components are equal and, when corrected for attenuation, are equal to unity so that each component is measuring the same variable, i.e.,

$$\frac{r_{jk}}{\sqrt{r_{jj} r_{kk}}} = 1 \tag{13}$$

and $r_{jj} = r_{kk}$, $r_{jk} = r_{jj}$, then equation (12) becomes

$$r_{yy} = 1 - \frac{1 - r_{jj}}{1 + (n-1) r_{jj}} = \frac{n r_{jj}}{1 + (n-1) r_{jj}}. \tag{14}$$

This is the familiar Spearman-Brown formula for estimating test reliability, seen here to be a special case of the reliability of a

weighted composite as given by equation (5) under the conditions that

$$W_j = W_k,\ \sigma_j = \sigma_k,\ r_{jj} = r_{kk},\ \text{and}\ \frac{r_{jk}}{\sqrt{r_{jj}\, r_{kk}}} = 1.$$

Returning to equation (12), it can be seen that the reliability of a composite of equally weighted standard scores increases with increasing average intercorrelation of the components. This conclusion has been stated before by Richardson and is implicit in the Richardson-Kuder treatment of reliability. The significance of its restatement here is that the same conclusion appears again as a lemma in the development of the Spearman-Brown formula, which has been considered by Kelley* as a measure of reliability of a different nature from the Richardson-Kuder reliability. The effect of the intercorrelations of the components is investigated further below.

Effect of Interrelationships of the Components. Certain effects of correlated vs. uncorrelated components on the reliability of the resulting composite can be deduced as special cases from equation (8), repeated here for ready reference.

$$r_{ww} = \frac{\sum_{j}^{n} w^2_j r_{jj} + 2\sum_{j=1}^{n-1}\sum_{k=2}^{n} w_j w_k r_{jk}}{\sum w^2_j + 2\sum_{j=1}^{n-1}\sum_{k=2}^{n} w_j w_k r_{jk}}. \qquad (8)$$

If the components are mutually uncorrelated, so that for any pair of components, j and k, $r_{jk} = 0$, equation (8) reduces to

$$r_{ww} = \frac{\sum w^2_j r_{jj}}{\sum w^2_j}. \qquad (15)$$

This is equivalent to the statement that, *for mutually uncorrelated components, the reliability of the weighted composite is a weighted mean of the reliabilities* the weight of each being w^2_j. For equal weights, this reduces to the mean reliability of the components, a value generally less than that of the most reliable of the sub-tests. In view of the tendency among some workers in the field to seek uncorrelated components for prediction, due to the dependence on the multiple regression concept,† this conclusion is of more than passing interest.

* Kelley, T. L., The Reliability Coefficient, *Psychometrika*, 1942, 7, 75-83.

† On this, cf. Richardson, M. W. "The combination of measures." In Horst, Paul, *The Prediction of Personal Adjustment*, Social Science Research Council, 1941, especially pp. 392 and 397.

The frequent loose statement of the Spearman-Brown formula, ignoring the assumptions following equation (14), as "The reliability of a test increases as its length increases," together with a loose application of the principle of multiple regression and its corollary that "Other things equal, $R_{y \cdot jk \cdots n}$ is a maximum if the intercorrelations, r_{jk} are a minimum," has led to the uncritical combination of unrelated measures to increase "validity" for some vague and ill-defined criterion, without regard for the effect on the reliability of the measure.

When, on the other hand, the intercorrelations are all equal to a given value, say $\bar{r}_{jk}$, equation (8) becomes

$$r_{yy} = \frac{\sum w^2_j r_{jj} + 2\bar{r}_{jk} \sum\sum w_j w_k}{\sum w^2_j + 2\bar{r}_{jk} \sum\sum w_j w_k}. \qquad (j < k) \qquad (16)$$

If, moreover, the reliabilities are also equal, this becomes

$$r_{yy} = \frac{r_{jj} \sum w^2_j + 2\bar{r}_{jk} \sum\sum w_j w_k}{\sum w^2_j + 2\bar{r}_{jk} \sum\sum w_j w_k}. \qquad (17)$$

If, however, the weights are equal, but the reliability unrestricted, equation (16) becomes (remembering $j < k$ and letting the mean of the n reliabilities be $\bar{r}_{jj}$),

$$r_{yy} = \frac{\bar{r}_{jj} + (n-1)\bar{r}_{jk}}{1 + (n-1)\bar{r}_{jk}}. \qquad (18)$$

The third possibility with regard to the intercorrelations of the components is that, when corrected for attenuation, they are equal to unity as in

$$r_{jk} = \sqrt{r_{jj}\, r_{kk}}. \qquad (13)$$

Under this condition, the expression for the reliability of a composite becomes of especial interest only if the reliabilities are equal. Then $r_{jk} = r_{jj}$ and equation (8) may be written

$$r_{yy} = r_{jj} \frac{\sum w^2_j + 2\sum\sum w_j w_k}{\sum w^2_j + 2 r_{jj} \sum\sum w_j w_k}. \qquad (19)$$

Since $r_{jj} \leq 1$, the denominator of the fraction is always less than the numerator and, hence, for equal reliabilities of tests all measuring the same factor, r_{yy} is greater than r_{jj}. If, as we have seen earlier, the weights are also equal, r_{yy} is given by the Spearman-Brown formula.

Weighting for maximum reliability. Given the algebraic statement of the reliability of the composite in terms of r_{jj}, r_{jk}, and w_j, in which the first two sets of values are given by the data, it would

appear simple to determine that set of weights yielding maximum reliability. The general approach is, of course, obvious—obtain the set of partial derivatives of equation (7) with respect to each w_j in turn, and solve the resulting set of homogeneous equations simultaneously, arbitrarily equating one of the w_j to unity. The actual solution for the general case, however, is far from simple.

Differentiating equation (7) with respect to each weight, w_p, in turn yields:

$$\frac{\partial r_{yy}}{\partial w_p} = -\frac{(2w_p - 2w_p r_{pp})(\sum w^2_j + 2\sum\sum w_j w_k r_{jk})}{(\sum w^2_j + 2\sum\sum w_j w_k r_{jk})^2} - \frac{(2w_p + 2\sum w_k r_{pk})(\sum w^2_j - \sum w^2_j r_{jj})}{(\sum w^2_j + 2\sum\sum w_j w_k r_{jk})^2}. \qquad (k \neq p) \quad (20)$$

If $\partial r_{yy}/\partial w_p = 0$, then the numerators of the two terms above are equal, thus:

$$w_p(1 - r_{pp})(\sum w^2_j + 2\sum\sum w_j w_k r_{jk}) = (w_p + \sum w_k r_{pk})(\sum w^2_j - \sum w^2_j r_{jj}). \qquad (21)$$

Equation (21) defines a set of n homogeneous simultaneous equations, since p may take each value from one through n. These equations can be solved only by imposing a further limitation.

It is convenient to define the weight of one variable, say the q-th component, as unity. Dividing each of the remaining $(n-1)$ equations of (21) by the equation resulting from equating $\frac{\partial r_{yy}}{\partial w_p}$ to zero gives

$$\frac{w_p(1 - r_{pp})}{w_q(1 - r_{qq})} = \frac{w_p + \sum_k^{n-1} w_k r_{pk}}{w_q + \sum_k^{n-1} w_k r_{qk}}, \qquad \begin{matrix}(k \neq p) \\ (k \neq q)\end{matrix} \qquad (22)$$

or, since $w_q = 1$,

$$w_p \frac{(1 - r_{pp})}{(1 - r_{qq})} = \frac{w_p + \sum_k^{n-1} w_k r_{pk}}{1 + \sum_k^{n-1} w_k r_{qk}}. \qquad (23)$$

Clearing of fractions yields

$$w_p(1 - r_{pp})(1 + \sum w_k r_{qk}) = w_p(1 - r_{qq}) + (1 - r_{qq})\sum w_k r_{pk}. \qquad (24)$$

Subtracting $w_p(1 - r_{qq})$ from both sides yields

$$w_p[(1 - r_{pp})(1 + \sum w_k r_{qk}) - (1 - r_{qq})] = (1 - r_{qq}) \sum w_k r_{pk} . \quad (25)$$

Expanding the bracketed term gives

$$w_p[1 - r_{pp} + \sum w_k r_{qk}(1 - r_{pp}) - 1 + r_{qq}] = (1 - r_{qq}) \sum w_k r_{pk} . \quad (26)$$

Solving explicitly for w_p,

$$w_p = \frac{\sum w_k r_{pk}(1 - r_{qq})}{(\sum w_k r_{qk})(1 - r_{pp}) + r_{qq} - r_{pp}} . \qquad \begin{matrix}(k \neq p) \\ (k \neq q)\end{matrix} \quad (27)$$

From equation (27) we can state that, for maximum reliability of the composite, the weight of a component is directly proportional to the weighted sum of its intercorrelations with the remaining components, and inversely proportional to its error variance.

The explicit statement of equation (27) is given below for the case of two variables, where variable 2 is taken as the q-th component.

$$\begin{aligned} w_1 &= \frac{w_2 r_{12}(1 - r_{22})}{w_2 r_{12}(1 - r_{11}) + r_{22} r_{11}} \\ w_2 &\equiv 1 , \end{aligned} \quad (28)$$

whence

$$w_1 = \frac{r_{12}(1 - r_{22})}{r_{12}(1 - r_{11}) + r_{22} - r_{11}} .$$

An exact solution to the equations for more than two variables is not apparent to the writer, although an iterative approximation should yield results. The difficulties of computation, however, render equation (27) of theoretical rather than practical value.

PSYCHOMETRIKA—VOL. 8, NO. 3
SEPTEMBER, 1943

MULTIPLE RECTILINEAR PREDICTION AND THE RESOLUTION INTO COMPONENTS: II

LOUIS GUTTMAN AND JOZEF COHEN*
CORNELL UNIVERSITY

Given a battery of n tests that has already been resolved into r orthogonal common factors and n unique factors, procedures are outlined for computing the following types of linear multiple regressions directly from the factor loadings: (i) the regression of any one test on the $n-1$ remaining tests; (ii) all the n different regressions of order $n-1$ for the n tests, computed simultaneously; (iii) the regression of any common factor on the n tests; (iv) the regressions of all the common factors on the n tests computed simultaneously; (v) the regression of any unique factor on the n tests; (vi) the regressions of all the unique factors on the n tests, computed simultaneously. Multiple and partial correlations are then determined by ordinary formulas from the regression coefficients. A worksheet with explicit instructions is provided, with a completely worked out example. Computing these regressions directly from the factor loadings is a labor-saving device, the efficiency of which increases as the number of tests increases. The amount of labor depends essentially on the number of common factors. This is in contrast to computations based on the original test intercorrelations, where the amount of labor increases more than proportionately as the number of tests increases. The procedures evaluate formulas developed in a previous paper (2). They are based essentially on a shortened way of computing the inverse of the test intercorrelation matrix by use of the factor loadings.

1. *Introduction*

Factor analysis and multiple correlation have at least three important contributions to be made by either one to the other:

(a) After a factor analysis has been performed, it is often possible to simplify the labor involved in computing multiple regressions.

(b) Factor analysis can give clues to aid in the selection of variables to use in regressions.

(c) Since it has been shown that common and unique factors are not perfectly predictable in general from a finite number of tests, the whole foundation of the theory of common and unique factors must be re-examined and reformulated if it is to have any meaning with respect to real phenomena.

These three points have been made, or are implicit, in a previous paper by Guttman (2). That paper is devoted to a large extent to

* The junior author worked out the numerical example and wrote explicit directions for the worksheet.

deriving formulas for point (a) above, both for predicting test scores and for predicting factor scores. These formulas are basic to discussions of points (b) and (c). The present paper is devoted to exhibiting explicit numerical procedures for the evaluation of these formulas. No knowledge is required here of Guttman's previous paper.

It is hoped that world events will be such that further papers, on points (b) and especially (c), can be prepared in the near future.

As has been pointed out in (2), the use of the formulas for point (a) does not depend upon the meaning of factors. It is a purely algebraic proposition which is universally legitimate as a mechanical means for computing ordinary multiple regressions.

2. *Prediction of Tests*

Conforming to terminology that has arisen from the fact that factor analysis was first developed in connection with psychological tests, we shall call any directly observed variable a *test*. The ordinary problem of multiple correlation is to predict a particular test from $n-1$ other tests. For a set of n tests, there are n different regressions of order $n-1$ (i.e., utilizing $n-1$ tests as predictors). The ordinary problem of partial correlation is to examine the correlation between two tests when $n-2$ other tests have been held constant. In the completely linear theory, to which we are restricting ourselves throughout this paper, for a set of n tests there are $\frac{n(n-1)}{2}$ different partial correlations, holding $n-2$ tests constant.

In standard methods for computing a multiple regression, the amount of labor increases far more than proportionately as the number of tests increases. In the fortunate case where it is possible to "factor" intercorrelations of tests into *common factor* loadings, where the number of common factors is much less than the number of tests, Guttman (2) has derived formulas that shorten the work considerably.* We shall give an explicit, compact, numerical method for evaluating these formulas, in terms of a numerical example. The computational procedures can be followed without any knowledge of Guttman's previous paper; they are handled routinely almost like an ordinary Doolittle arrangement. In the present method, the amount of labor does *not* increase much proportionately as the number of tests increases. The amount of labor depends most essentially on the number of common factors, rather than on the number of tests.

By the present method, it is relatively easy to compute the re-

* Much of the same ground has also been covered, with a rather different emphasis, by Dwyer in (1).

gression coefficients for the regression of any one test on the $n-1$ remaining tests, no matter how many tests there are, provided the number of common factors is relatively small. It is also relatively easy to compute all the regression coefficients for *all the n different possible regressions* of order $n-1$ for the n tests. As is well known, when all the regression coefficients are given, it is a simple matter to compute all the multiple and partial *correlation* coefficients.

The numerical example in §5 below shows how to compute any particular regression, and how to compute simultaneously all the possible regressions of order $n-1$.

3. *Prediction of Factors*

Using a theory of common and unique factors to "factor" the intercorrelations of n tests introduces, in general, more factors than there are tests. In general, there are n unique factors—one for each test—and, say, r common factors, making a total of $n+r$ factors. The "factoring" process defines the correlations of the factors with the tests, these correlations being called the *factor loadings*.* However, such a "factoring" process *cannot* define the score of a person on a factor, as long as n is finite. This follows from the fact that $n+r$ linearly independent variables cannot be expressed as linear functions of n variables. The only direct observations on hand are the scores of persons on the n tests, so that scores on the factors must be obtained somehow as functions of the observed test scores. But there are too many factors to be able to do this exactly. Hence the factor scores must be *estimated* or predicted from the test scores. This requires the multiple *regression equation* of each factor on the n tests, which we *can* compute exactly because only correlation coefficients are needed in the computations. We do know the intercorrelations between the n tests, and the correlations of the factors with the tests.

These regressions are, then, of order n, since the n tests are used as predictors. However, there are simplified formulas again which make the labor involved depend essentially on the number of common factors, rather than on the number of tests. In the numerical example below, it is shown how to compute the regression of any particular common factor on the n tests, and of any particular unique factor on the n tests. It is also shown how to compute all the regressions for the factors simultaneously.

Again, once the regression coefficients have been computed, it is a simple matter to ascertain the multiple correlation coefficients.

* The process also defines the correlations *between* the factors. We deal here only with uncorrelated factors. It is not difficult to extend the present procedure to the case of correlated factors.

4. *The Matric Formulas*

This section may be omitted by non-mathematical readers. It is a summary of the formulas derived in Guttman's previous paper [see (2), pp. 91-92].

The matrix of the intercorrelations of the tests is denoted by $\mathbf{R}$. It is of order $n \times n$. The $n \times r$ matrix of common factor loadings is denoted by $\mathbf{A}$. The $n \times n$ diagonal matrix of unique factor loadings is denoted by $\mathbf{U}$. From the fundamental theorem of factor analysis,

$$\mathbf{R} = \mathbf{A}\mathbf{A}' + \mathbf{U}^2,$$

where the prime denotes the transposed matrix. All regression coefficients for predicting tests from $n-1$ other tests are proportional to the elements of $\mathbf{R}^{-1}$. Hence, the computation of $\mathbf{R}^{-1}$ gives essentially all possible regressions of order $n-1$. The computational procedures using factor loadings are merely methods of simplifying the calculation of $\mathbf{R}^{-1}$, or of $\mathbf{R}^{-1}$ multiplied by given matrices.

Let $\mathbf{Q} = \mathbf{I} + \mathbf{A}'\mathbf{U}^{-2}\mathbf{A}$. Then $\mathbf{Q}$ is non-singular and of order $r \times r$. Guttman shows that

$$\mathbf{R}^{-1} = \mathbf{U}^{-2} - \mathbf{U}^{-2}\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}'\mathbf{U}^{-2}. \qquad (1)$$

The only difficult computation in the right member is to compute the inverse of the $r \times r$ matrix $\mathbf{Q}$. In Table 5 below, the entire right member of (1) is computed as a single process.

The regression weights for predicting the r common factors from the n tests are the elements of

$$\mathbf{A}'\mathbf{R}^{-1} = \mathbf{Q}^{-1}\mathbf{A}'\mathbf{U}^{-2}. \qquad (2)$$

Ledermann (5) has given a numerical model for computing the *right* member of (2), using Aitken's pivotal condensation. In the numerical example below, we actually get the regression for the rth common factor in essentially this manner, using instead the Doolittle arrangement. A Doolittle arrangement can of course be set up which would be equivalent to Ledermann's procedure, but this is not a purpose of the present paper. Instead, we utilize the fact that $\mathbf{R}^{-1}$ has been computed from the right member of (1). Then the *left* member of (2) is easily computed by direct multiplication.

Similarly, the regression weights for predicting the unique factors from the tests are the elements of

$$\mathbf{U}\mathbf{R}^{-1}. \qquad (3)$$

Once $\mathbf{R}^{-1}$ has been computed, according to the right member of (1), (3) is immediately obtainable.

5. *Preparing the Worksheet*

As a numerical example, we shall take a fictitious problem of Holzinger [(4), pp. 32-36], which was also used by Ledermann (5). In Table 1 are the intercorrelations of Holzinger's seven fictitious tests, and in Table 2 are the factor loadings on a particular set of factors. In this example, there are three common factors as against seven tests, so that the saving in labor will not be too considerable by the present method over some other method operating directly on the correlation coefficients. The example is only to illustrate the method, and it will be apparent how considerable is the labor saved when there are, say, fifteen or twenty or more tests and only three or four common factors, or fifty tests with seven or eight common factors. Even for six or seven tests, if there were but one common factor, there would be little work at all in the present method.

In all the computations below, the work was actually carried to three decimals, but the results are presented here only to two places.

In the bottom section of Table 2, the first line contains the communality for each test, i.e., the sum of the squares of the common factor loadings, denoted by h^2. The uniqueness of a test is $u^2 = 1 - h^2$, as recorded in the second line. The reciprocal of each uniqueness is recorded in the third line, to be used for computing Table 3. The loading of a test on its unique factor is u, the (positive) square root of the uniqueness, as recorded in the last line.

Table 3 is obtained by multiplying each column of the first section of Table 2 by its corresponding value of $1/u^2$. This augments the common factor loadings of a test according to the uniqueness of the test.

Table 4 is actually a symmetrical table, so that the entries below the principal diagonal have been omitted for convenience. An entry in row i and column j of Table 4 is obtained by taking each entry in row i of Table 2, multiplying it by the corresponding element of row j in Table 3, and summing the n products (n is the number of tests). For the *diagonal elements*, unity is added to the sum. Thus, the entry in row I, column II of Table 4, is

$$(.7)(1.92) + (.6)(2.14) + (.5)(7.27) + (.7)(0) + (.6)(0) + (.4)(0) + (.8)(0) = 6.27.$$

The diagonal element in the third row and column is

$$0 + 0 + 0 + (.6)(4.00) + (.6)(2.14) + (.8)(4.00) + 0 = 6.89$$

plus unity.

All the requisite entries for the worksheet, Table 5, are now on hand. The instructions for the worksheet are self-explanatory, being rather similar to those for a Doolittle arrangement.

Table 6 actually is seven (in general, n) lines of the worksheet. Row 1 of Table 6 is the sum of four (in general, $r + 1$) lines of Table 5: 12 + 13 + 14 + 15; row 2 is 16 + 17 + 18 + 19; etc. The entries below the diagonal are filled in by symmetry. The complete table is the inverse of the correlation matrix of Table 1.

Table 6—and the worksheet—can be checked as follows: the sum of the products of the entries in row i of Table 1 by the corresponding entries in row j of Table 6 should be zero (within rounding errors) if $i \neq j$, and it should be unity (within rounding errors) if $i = j$. (In the matric formulation, this follows from the symmetry of $\mathbf{R}$ and the fact that $\mathbf{RR}^{-1} = \mathbf{I}$.)

6. *When the Entire Worksheet is Not Needed*

The entire worksheet above is not needed for many problems. If it is desired to find the regression of only a particular *common factor* on the n tests, the whole bottom section of the worksheet can be omitted. For example, on the present worksheet, we would need only the first eleven lines. The last seven entries on line 11 are actually the *negatives* of the regression coefficients of the tests for predicting common factor III (see §8 below). Had we wanted, instead, to predict common factor II, say, we would have prepared the worksheet with the rows and columns for common factor II in the positions where those for common factor III are now.

It would be very easy to modify the above worksheet to give directly the regressions for *all* the common factors on the tests. This is done, for example, by Ledermann (5) using an Aitken arrangement. We shall not take the space here to give the equivalent Doolittle arrangement. What we shall do here is take advantage of the fact that the entire worksheet *will* be used for many problems, so that we shall use below a procedure for getting the regressions for all the common factors *after* the entire worksheet is completed.

If it is desired to compute the regression of only a particular *test* and/or a particular *unique* factor, the entire worksheet is not necessary. On the present worksheet, if we wished to compute the regression of the first test and/or the first unique factor, we would need only the first fifteen lines on the worksheet, from which we get the first line of Table 6. As is indicated in computing Tables 8 and 12 below, the entries in the first line of Table 6 are proportional to the regression coefficients for predicting the first test and the first unique

factor. Again, if instead we had wished to predict only the jth test and/or unique factor, we would have arranged the rows and columns of the jth test and unique factor where those for the first test and unique factor are now on the worksheet.

The entire worksheet is necessary for computing simultaneously the regressions for all tests and/or all unique factors.

7. *The Complete Set of Regressions for Tests*

To aid in computing the multiple and partial coefficients for predicting tests, Table 7 is set up. The first row consists of the entries in the principal diagonal of Table 6, denoted by d. The second and third rows, as indicated in Table 7, are the reciprocal and square root of the reciprocal, respectively, of the corresponding element in the first row.

The left-hand section of Table 8 is got from Table 6 by multiplying each row in Table 6 by the *negative* of the corresponding element in the second row of Table 7 (namely, $-1/d$). For example, the third row of Table 8 is —.31 times the third row of Table 6.

The diagonal entries of Table 8 should all be —1.00 if computed; and they are omitted, since they are not regression coefficients but merely check the work. The jth row of Table 8 gives the $n-1$ regression coefficients for predicting the jth test from the $n-1$ remaining tests (when all tests are in standard form).

The *index of multiple correlation* is the term used here to denote the square of the multiple correlation coefficient.* As is well known, if r_{jk} is the correlation between variables j and k, and if b_{jk} is the regression coefficient of variable k in the regression of variable j on, say, m variables, then the index of multiple correlation of variable j on these variables can be computed from the formula

$$r_{j1}b_{j1} + r_{j2}b_{j2} + \cdots + r_{jm}b_{jm}$$

(assuming all the variables are in standard form). The multiple correlation coefficient is then, of course, the positive square root of this index.

Thus, the index of multiple correlation of the third test in Table 8, for example, is computed by multiplying each element in the third row of Table 1—omitting the diagonal entry—by the corresponding element in the third row of Table 8, and adding the $n-1$ products:

* This is a far better term to use than "index of determination," which gives misleading notions about causation. A further contribution of factor analysis to multiple correlation, which might be added as a fourth point to those at the beginning of this paper, is to afford a good device for showing how *meaningless* multiple correlation may be from the point of view of a causal analysis. See Section 20 of (3), pp. 289-291.

$$(.75)(.44) + (.78)(.52) + (.35)(-.05) + (.30)(-.02)$$
$$+ (.20)(.02) + (.40)(-.06) = .69.$$

This result is recorded in the next to the last column of Table 8. In the last column is recorded the multiple correlation coefficient, .83, which is the positive square root of .69.

To obtain the complete set of *partial* correlation coefficients, Table 9 is first computed by multiplying each row of Table 6 by the negative of the corresponding element in the third row of Table 7, namely, $-\sqrt{1/d}$. For example, the third row of Table 9 is $-.56$ times the third row of Table 6.

Table 10 is then obtained by multiplying each *column* of Table 9 by the corresponding entry in the third row of Table 7, namely, $\sqrt{1/d}$. For example, the fourth column of Table 10 is .50 times the fourth column of Table 9. Except for rounding errors, Table 10 is symmetric; and it gives all the partial correlation coefficients of order $n-2$. The diagonal entries are omitted. They should be -1.00 if computed as a check on the work.

8. *The Complete Set of Regressions for Factors*

Each row in Table 11 contains the n regression coefficients for predicting the indicated common factor from the n tests (when all the variables are in standard form). The regression coefficient in the ith row and jth column is obtained by multiplying each element in the ith row of (the top section of) Table 2 by the corresponding element in the jth row of Table 6 and adding the n products. For example, the regression coefficient for the fourth test for predicting the first common factor is

$$(.7)(-.47) + (.6)(-.23) + (.5)(.17) + (.7)(4.02)$$
$$+ (.6)(-1.29) + (.4)(-1.72) + (.8)(-.77) = .35.$$

Notice that the last row of Table 11 is the same, except for sign, as the last row of the top section of Table 5 (line 11). This is in accordance with what was pointed out in §6, that the entire worksheet is not needed if only a particular common factor is to be predicted. This also affords a numerical check on the work.

The indexes of multiple correlation are computed by the general formula exhibited above in §7. This time the variable predicted is a common factor, and the correlation coefficients are the common factor loadings. The index for the jth common factor is obtained by multiplying each element in the jth row of (the top section) of Table 2 by the corresponding element in the jth row of Table 11 and summing the products. This time, the diagonal entries in Table 6 are included

in the calculations, since all tests are used as predictors.

Again, the coefficient of multiple correlation is the positive square root of the index of correlation.

The computations for predicting the unique factors are rather simple. The regression coefficients in Table 12 are obtained by multiplying each row of Table 6 by the corresponding entry in the seventh row of Table 2, namely u, the unique loading of the test containing the unique factor. For example, the fifth row in Table 12 is .53 times the fifth row of Table 6.

The indexes of multiple correlation for the unique factors are again obtained by the general formula exhibited in §7 above. This time, the variable predicted is a unique factor, and the correlation coefficients are the correlations of the unique factor with the n tests. But these are all zero by the definition of a unique factor, except for the test containing the unique factor; so that $n-1$ of the terms in the sum vanish, leaving only the term from the test containing the unique factor. Thus, the index of multiple correlation of the jth unique factor on the n tests is simply the unique loading of the jth test times the regression coefficient of the jth test. Therefore, the column of indexes in Table 12 is obtained simply by multiplying each diagonal element in the table by the corresponding element in the last row of Table 2 (namely, u). For example, the index for the second unique factor is

$$(1.58)\,(.53) = .84\,.$$

Alternatively, and with more accuracy with respect to decimals, these same indexes could be obtained by multiplying the diagonal elements of Table 6 by the corresponding elements in fifth row of Table 2 (namely, u^2, the uniqueness). Thus the index for the second unique factor is

$$(2.98)\,(.28) = .83\,.$$

Again, the multiple correlation coefficient is the positive square root of the correlation index.

TABLE 1
The Intercorrelations of the Tests
(The Matrix **R**)

Test	Test 1	2	3	4	5	6	7
1	1.00	.72	.75	.49	.42	.28	.56
2	.72	1.00	.78	.42	.36	.24	.48
3	.75	.78	1.00	.35	.30	.20	.40
4	.49	.42	.35	1.00	.78	.76	.56
5	.42	.36	.30	.78	1.00	.72	.48
6	.28	.24	.20	.76	.72	1.00	.32
7	.56	.48	.40	.56	.48	.32	1.00

TABLE 2
Factor Loadings
(The Matrix **A'**)

Common Factor	Test 1	2	3	4	5	6	7
I	.7	.6	.5	.7	.6	.4	.8
II	.5	.6	.8				
III				.6	.6	.8	
Communality (h^2)	.74	.72	.89	.85	.72	.80	.64
Uniqueness (u^2)	.26	.28	.11	.15	.28	.20	.36
$1/u^2$	3.85	3.57	9.09	6.67	3.57	5.00	2.78
Unique Loading (u)	.51	.53	.33	.39	.53	.45	.60

TABLE 3
Augmented Common Factor Loadings
(The Matrix $\mathbf{A'U^{-2}}$)

Common Factor	Test 1	2	3	4	5	6	7
I	2.70	2.14	4.54	4.67	2.14	2.00	2.22
II	1.92	2.14	7.27				
III				4.00	2.14	4.00	

TABLE 4
The Matrix **Q**
($\mathbf{Q = I + A'U^{-2}A}$)

Common Factor	Common Factor I	II	III
I	13.57	6.27	5.69
II		9.06	
III			7.89

DIRECTIONS FOR FILLING OUT THE WORKSHEET

The following are examples of how to read the symbols used in the directions below:

22.5 refers to the entry in line 22, column 5 of the worksheet
6.III refers to the entry in line 6, column III
22.5 - 22.7 refers to the entries 22.5, 22.6, 22.7
22 refers to all entries in line 22

Line:

1. Insert in 1.I-1.III the first row of Table 4; insert in 1.1-1.7 the first row of Table 3
2. Divide 1 by the negative of 1.I
3. Insert in 3.II-3.III the second row of Table 4; insert in 3.1-3.7 the second row of Table 3
4. Multiply 1.II-1.7 by 2.II
5. 3 + 4
6. Divide 5 by the negative of 5.II
7. Insert in 7.III the third row of Table 4; insert in 7.1-7.7 the third row of Table 3
8. Multiply 1.III-1.7 by 2.III
9. Multiply 5.III-5.7 by 6.III
10. 7 + 8 + 9
11. Divide 10 by the negative of 10.III
12. Insert in 12.1 the $\frac{1}{u^2}$ for test 1 from Table 2
13. Multiply 1.1-1.7 by 2.1
14. Multiply 5.1-5.7 by 6.1
15. Multiply 10.1-10.7 by 11.1
16. Insert in 16.2 the $\frac{1}{u^2}$ for test 2 from Table 2
17. Multiply 1.2-1.7 by 2.2
18. Multiply 5.2-5.7 by 6.2
19. Multiply 10.2-10.7 by 11.2
20. Insert in 20.3 the $\frac{1}{u^2}$ for test 3 from Table 2
21. Multiply 1.3-1.7 by 2.3
22. Multiply 5.3-5.7 by 6.3
23. Multiply 10.3-10.7 by 11.3
24. Insert in 24.4 the $\frac{1}{u^2}$ for test 4 from Table 2

25 Multiply 1.4-1.7 by 2.4

26. Multiply 5.4-5.7 by 6.4
27. Multiply 10.4-10.7 by 11.4
28. Insert in 28.5 the $\frac{1}{u^2}$ for test 5 from Table 2
29. Multiply 1.5-1.7 by 2.5
30. Multiply 5.5-5.7 by 6.5
31. Multiply 10.5-10.7 by 11.5

32. Insert in 32.6 the $\frac{1}{u^2}$ for test 6 from Table 2
33. Multiply 1.6-1.7 by 2.6
34. Multiply 5.6-5.7 by 6.6
35. Multiply 10.6-10.7 by 11.6
36. Insert in 36.7 the $\frac{1}{u^2}$ for test 7 from Table 2
37. Multiply 1.7 by 2.7
38. Multiply 5.7 by 6.7
39. Multiply 10.7 by 11.7

TABLE 5

Worksheet for Computing the Inverse of the Correlation Matrix

Line	I	II	III	1	2	3	4	5	6	7
1	13.57	6.27	5.69	2.70	2.14	4.54	4.67	2.14	2.00	2.22
2	−1.00	−.46	−.42	−.20	−.16	−.34	−.34	−.16	−.15	−.16
3		9.06	0.00	1.92	2.14	7.27				
4		−2.90	−2.63	−1.24	−.99	−2.10	−2.16	−.99	−.92	−1.03
5		6.17	−2.63	.68	1.15	5.17	−2.16	−.99	−.92	−1.03
6		−1.00	.43	−.11	−.19	−.84	.35	.16	.15	.17
7			7.89				4.00	2.14	4.00	
8			−2.38	−1.13	−.90	−1.90	−1.96	−.90	−.84	−.93
9			−1.12	.29	.49	2.20	−.92	−.42	−.39	−.44
10			4.39	−.84	−.41	.30	1.13	.82	2.77	−1.37
11			−1.00	.19	.09	−.07	−.26	−.19	−.63	.31
12				3.85						
13				−.53	−.42	−.90	−.92	−.42	−.40	−.44
14				−.08	−.13	−.57	.24	.11	.10	.11
15				−.16	−.08	.06	.22	.16	.53	−.26
16					3.57					
17					−.34	−.72	−.74	−.34	−.32	−.35
18					−.22	−.97	.40	.19	.17	.19
19					−.04	.03	.11	.08	.26	−.13
20						9.09				
21						−1.52	−1.56	−.72	−.67	−.74
22						−4.34	1.81	.83	.78	.86
23						−.02	−.08	−.06	−.19	.09
24							6.67			
25							−1.61	−.74	−.69	−.76
26							−.75	−.35	−.32	−.36
27							−.29	−.21	−.71	.35
28								3.57		
29								−.34	−.32	−.35
30								−.16	−.15	−.16
31								−.16	−.52	.26
32									5.00	
33									−.29	−.33
34									−.14	−.15
35									−1.75	.86
36										2.78
37										−.36
38										−.17
39										−.43

TABLE 6

The Inverse of the Correlation Matrix

$(\mathbf{R}^{-1} = \mathbf{U}^{-2} - \mathbf{U}^{-2}\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}'\mathbf{U}^{-2})$

Test	Test 1	2	3	4	5	6	7
1	3.08	−.63	−1.41	−.47	−.16	.24	−.59
2	−.63	2.98	−1.66	−.23	−.08	.11	−.29
3	−1.41	−1.66	3.21	.17	.06	−.08	.21
4	−.47	−.23	.17	4.02	−1.29	−1.72	−.77
5	−.16	−.08	.06	−1.29	2.92	−.98	−.26
6	.24	.11	−.08	−1.72	−.98	2.82	.38
7	−.59	−.29	.21	−.77	−.26	.38	1.82

TABLE 7

Principal Diagonal Entries*

(Diagonal Matrices $\mathbf{D}$, $\mathbf{D}^{-1}$, $\mathbf{D}^{-1}$)

	Test 1	2	3	4	5	6	7
d	3.08	2.98	3.21	4.02	2.92	2.82	1.82
$\frac{1}{d}$	.32	.34	.31	.25	.34	.35	.55
$\frac{1}{\sqrt{d}}$	.57	.58	.56	.50	.58	.60	.74

* Each row of this table may be thought of as defining a diagonal matrix, the entries in each row of the table being the principal diagonal elements of the matrices.

TABLE 8

Multiple Regression Coefficients for Predicting the Tests

(The Matrix $-\mathbf{D}^{-1}\mathbf{R}^{-1}$)

Predicted Test	Test (Predictor) 1	2	3	4	5	6	7	Multiple Correlation Index	Coefficient
1		.20	.46	.15	.05	−.08	.19	.67	.82
2	.21		.56	.08	.03	−.04	.10	.67	.82
3	.44	.52		−.05	−.02	.02	−.06	.69	.83
4	.12	.06	−.04		.32	.43	.19	.75	.86
5	.05	.03	−.02	.44		.34	.09	.66	.81
6	−.08	−.04	.03	.61	.35		−.13	.64	.80
7	.32	.16	−.12	.42	.14	−.21		.44	.67

TABLE 9
(The Matrix $-\mathbf{D}^{-1}\mathbf{R}^{-1}$)

Test	Test 1	2	3	4	5	6	7
1	−1.75	.36	.80	.27	.09	−.14	.34
2	.36	−1.73	.96	.13	.05	−.06	.17
3	.79	.93	−1.79	−.09	−.03	.04	−.12
4	.23	.11	−.08	−2.01	.64	.86	.38
5	.09	.05	−.04	.75	−1.71	.57	.15
6	−.14	−.07	.05	1.02	.58	−1.68	−.23
7	.44	.21	−.16	.57	.19	−.28	−1.35

TABLE 10
Partial Correlation Coefficients
(The Matrix $-\mathbf{D}^{-1}\mathbf{R}^{-1}\mathbf{D}^{-1}$)

Test	Test 1	2	3	4	5	6	7
1		.21	.45	.13	.05	−.08	.25
2	.21		.54	.06	.03	−.04	.12
3	.45	.54		−.04	−.02	.03	−.09
4	.13	.06	−.04		.37	.51	.28
5	.05	.03	−.02	.37		.34	.11
6	−.08	−.04	.03	.51	.34		−.17
7	.25	.12	−.09	.28	.11	−.17	

TABLE 11
Multiple Regression Coefficients for Predicting the Common Factors
(The Matrix $\mathbf{A}'\mathbf{R}^{-1}$)

Common Factor	Test 1	2	3	4	5	6	7	Multiple Correlation Index	Coefficient
I	.27	.12	−.09	.35	.12	−.17	.43	.81	.90
II	.03	.15	.87	−.24	−.08	.12	−.30	.80	.89
III	−.19	−.10	.07	.26	.19	.64	−.31	.78	.88

TABLE 12
Multiple Regression Coefficients for Predicting the Unique Factors
(The Matrix $\mathbf{UR}^{-1}$)

Unique Factor	1	2	3	4	5	6	7	Multiple Correlation Index	Multiple Correlation Coefficient
1	1.57	−.32	−.72	−.24	−.08	.12	−.30	.80	.89
2	−.33	1.58	−.88	−.12	−.04	.06	−.15	.84	.92
3	−.47	−.55	1.06	.06	.02	−.03	.07	.35	.59
4	−.18	−.09	.07	1.56	−.50	−.67	−.30	.41	.64
5	−.08	−.04	.03	−.68	1.55	−.52	−.14	.82	.91
6	.11	.05	−.04	−.77	−.44	1.26	.17	.57	.76
7	−.35	−.17	.13	−.46	−.16	.23	1.09	.65	.81

REFERENCES

1. Dwyer, Paul. The evaluation of multiple and partial correlation coefficients from the factorial matrix. *Psychometrika*, 1940, 5, 211-232.
2. Guttman, Louis. Multiple rectilinear prediction and the resolution into components. *Psychometrika*, 1940, 5, 75-99.
3. Guttman, Louis. An outline of the statistical theory of prediction. Supplementary study B-1 in Paul Horst et al., *The prediction of personal adjustment*, 1941, Social Science Research Council. Pp. 250-312.
4. Holzinger, Karl J. *Student manual of factor analysis.* Department of Education, University of Chicago, 1937.
5. Ledermann, Walter. On a shortened method of estimation of mental factors by regression. *Psychometrika*, 1939, 4. 109-116.

SEPTEMBER, 1943
PSYCHOMETRIKA—VOL. 8, NO. 3

A PREFERENTIAL MATCHING PROBLEM

LT. J. A. GREENWOOD, U. S. N. R.
BUREAU OF AERONAUTICS, NAVY DEPARTMENT, WASHINGTON, D. C.

A matching method proposed by Dr. C. E. Stuart is presented in some detail and the essentials for a test of significance are derived. This method differs from the older matching methods in that partial credit is allowed for a near miss. A slight variation of the method permits the matching of one item with M sets of n traits.

The matching method provides a means of applying statistical treatment to observations which cannot easily be classified according to fixed criteria. A simple form of the matching problem is the following: Suppose five sets of identical twins provide handwriting specimens. Five specimens, consisting of one specimen from each pair, are placed before a judge. He is asked to match the other five specimens to the first set according to degree of similarity. His resulting five similar pairs are then checked for agreement with the original twin pairing. If a consistently greater-than-chance agreement is noted, then it is possible to conclude safely that the handwriting of identical twins is similar even though there exist no fixed objective criteria of similarity.

In recent years there has been considerable literature on the subject of matching techniques similar to the above. As far as I have ascertained, they have dealt only with the numbers of perfect correspondences. Stuart (1) has recently proposed a technique in which each of the comparing items is "preferentially matched" or rated against each of the compared ones, depending on their closeness of "similarity."

This paper presents the mathematics of the latter situation and in addition presents a new aspect of the method. It is divided into cases (i), (ii), and (iii), of which the first two represent a formal description and derivation of the mathematics of Stuart's method. A brief description of the matching method under discussion follows.

Let two sets of items to be matched be A_i and B_i, $i = 1, 2, \cdots, n$. A_1 is compared to each of the B_i and the one which it most nearly "resembles" is given the number n,* the one next to it in "resem-

* Stuart gave it the value of 1. The mean and variance of the score are the same for the two ways of numbering.

blance" is given the number $n-1$, and so on to 1. This procedure is done with each A_i. Suppose A_1 really paired with B_1. Then with perfect recognition of the true association, B_1 would have been allocated n. However, any number from 1 to n may actually have been given it. The numbers thus allocated to the true correspondences of the A_i are summed as the test score.

It is clear that the set of assignments can be displayed in an n-square matrix where the number in the i-th row and j-th column is that assigned when A_i is compared with B_j. The test score is then the sum of the numbers on the principal diagonal,* providing A_i and B_i are true correspondences for $i = 1, 2, \cdots, n$. Thus a matrix with $n = 4$ and test score of 12 is illustrated by the accompanying diagram.

	B_1	B_2	B_3	B_4
A_1	4	1	3	2
A_2	3	1	2	4
A_3	3	1	4	2
A_4	2	4	1	3

Two cases are distinguished in practice. Case (i): The allocations in any row are considered independent of all the other rows except inasmuch as actual similarities influence the decisions. Case (ii): Each column (and of course each row) must contain an n. This necessitates remembering which B_i has already been allocated a value n, and not thereafter assigning it an n. Something between case (i) and case (ii) appears to take place in actual experiments. That is, the judges appear to be influenced by their previous allocations of "firsts" - - those receiving n - - but the difficult memory feat involved seems to preclude its effective extension to "seconds," etc.

In both cases (i) and (ii), the null hypothesis assumes that a priori each matching of an A_i with a B_j $(i, j = 1, 2, \cdots, n)$ has the same probability of an $n, n-1, \cdots,$ or 1 assignment. This is obviously not to say that the assignments are independent.

Case (i)

On a chance basis the probability of a score of r is the coefficient of x^r in

$$\phi(x) = \left(\frac{x + x^2 + \cdots + x^n}{n}\right)^n.$$

* That is, from the upper left-hand corner to the lower right-hand corner.

Therefore, the j-th factorial moment of r is $K_j = \phi^{[j]}(1)$, the j-th derivative of $\phi(x)$ evaluated at $x = 1$. In order to obtain the moments about the mean, μ_i, $i = 2, 3, 4$, it is thus necessary to obtain the first four derivatives of $\phi(x)$.* This is facilitated by carrying out the differentiation of $\phi(x)$ in terms of $A(x) = x + x^2 + \cdots + x^n$ and its derivatives. The straightforward differentiation is omitted, but the essential values used are

$$A^{[i]}(1) = \frac{(n+1)^{(i+1)}}{i+1}.\dagger$$

Substituting these derivatives evaluated at $x = 1$ in the formulas connecting factorial moments K_j and ordinary central moments μ_i, we obtain the following:

$$\text{mean} = K_1 = n(n+1)/2,$$

$$\text{variance} = \mu_2 = K_2 + K_1 - K_1^2 = (n+1)^{(3)}/12,$$

$$\mu_3 = K_3 + \mu_2(3 - 3K_1) - (K_1^3 - 3K_1^2 + 2K_1) = 0,$$

$$\begin{aligned}\mu_4 &= K_4 + \mu_3(6 - 4K_1) - \mu_2(6K_1^2 - 18K_1 + 11) \\ &\quad + (6K_1^3 - K_1^4 - 11K_1^2 + 6K_1) \\ &= \frac{(n+1)^{(3)}}{240}(5n^3 - 2n^2 - 5n - 2),\end{aligned}$$

$$\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = 0, \quad \alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{3(5n^3 - 2n^2 - 5n - 2)}{5n(n^2 - 1)}$$

$$\lim_{n\to\infty} \alpha_4 = 3.$$

Case (ii)

Let

$$\underline{|r} = \sum_{i=0}^{r} (-1)^i \binom{r}{i} (r-i)!, \quad \underline{|0} = 1, \quad (2).$$

Let $M_q^m = \binom{m}{q} \underline{|m-q}$ be the number of terms in a determinant of order m, each term having q elements, only, of the principal diagonal. The probability of j n's on the principal diagonal of the matrix of

* We note that for case (i) the desired moments could be obtained ultimately from the moments of the score of a single cell of the matrix.

† Define $a^{(b)} = a(a-1)\cdots(a-b+1)$, b a positive integer.

numbers is then $M_j^n/n!$. The probability generating function of the total score on the principal diagonal is $\Psi(x)$ where

$$n!\Psi(x) = \sum_{i=0}^{n} M_i^n x^{ni} \left(\frac{x + x^2 + \cdots + x^{n-1}}{n-1}\right)^{n-1i}.$$

Again the mean score K_1 equals $\Psi^{[1]}(1)$ and $K_2 = \Psi^{[2]}(1)$.

In order to maintain the upper limit n on the summation sign throughout the successive differentiations, it is sufficient to bound x away from zero in a neighborhood about 1. The following values are also needed.

$$\sum_{i=0}^{n} M_i^n = n! \text{ since } \Psi(1) = 1;$$

$$i\, M_i^n = n\, M_{i-1}^{n-1} \text{ so } \sum_{i=0}^{n} i\, M_i^n = n!\ ;$$

$$\sum_{i=0}^{n} i^2 M_i^n = 2(n!).$$

Carrying through the differentiation in terms of $\alpha(x) = (x + x^2 + \cdots + x^{n-1})/(n-1)$, from which $\alpha(1) = 1$, $\alpha^{[1]}(1) = n/2$, $\alpha^{[2]}(1) = n(n-2)/3$, one readily obtains $K_1 = n(n+1)/2$ as in case (i).

After substituting the derivatives of $\alpha(x)$ in $\Psi^{[2]}(x)$, all evaluated at $x = 1$, there results $K_2 = \Psi^{[2]}(1) = \frac{n}{12}(3n^3 + 7n^2 - 3n - 4)$,

variance $= \mu_2 = K_2 + K_1 - K_1^2 = \frac{n}{12}(n^2 + 2) = \left(1 + \frac{3}{n_2 - 1}\right)$ times μ_2 of case (i).

To obtain the third and fourth moments of case (ii) by this method of attack would be considerable of an exercise. If it is desired to evaluate the sum of results of a repetition of such an experiment as encompassed by either case (i) or case (ii), one need only apply the usual formulas for the moments of a mean.

As an illustration of this technique, one attempts to identify authors of handwriting specimens with character sketches of each. Let

n = number of specimens (or sketches),
A_i = handwriting specimen of i-th author,
B_j = character sketch of j-th author.

In the experiment to test whether or not to a given judge the handwriting specimens exhibit significantly identifying properties in common with the true associated character sketches, all identifying marks should be removed from all the A's (or the B's) and a code substituted before handing over to the judge. The procedure described previously should then be carried out by the judge and the score totaled by the

experimenter, after decoding so that the true correspondences of the A's can be identified.

Now it might well happen that A_1, say, suggested both B_3 and B_5 and that B_5 is the true associate of A_1. The usual "correct matching" procedure depends for *any* value of the matching score upon deciding correctly which of the two it should be, and further, if it is missed, it affects adversely the possibility of correctly associating the remaining ones. (This is the method where in effect the B's are lined up and the A's opposite them in one-to-one correspondence and correct pairings counted, for the score.) On the other hand, in the presently suggested method, at worst B_5 is allocated $n - 1$ and B_3 the value n. Thus it contributes $n - 1$ to the total score and has no effect in case (i) and little effect in case (ii) on the remaining matchings. *This would appear to indicate the essential advantage of the proposed method.*

In order to test the null hypothesis, the difference between the observed score and the theoretical value $n(n + 1)/2$ is divided by the standard deviation of case (i) or case (ii) (whichever the experimenter decides is more applicable) and the normal tables entered for an approximate probability. In the illustration at the beginning of the article $n = 4$, score $= 12$, theoretical mean $= 10$, and standard deviation [assuming case (ii)] is $\sqrt{6}$. While for only one sample n is too small to admit of accurate evaluation on the assumption of normality,* the formal procedure gives a critical ratio of $(11.5 - 10)/\sqrt{6} = 0.61$ with a probability of .3, roughly, of obtaining a score as large as 12.†.

Case (iii)

A different situation is involved in the following illustration. Choose M subjects (of comparable age and physical condition) and list n selected trait descriptions of each, as: measure of initiative, sample of handwriting, reaction time on a certain test, etc. Suppose the question is whether or not a photograph of a selected person of

* Even in the more favorable case (i), it appears that for a reasonably good evaluation of the single sample by normal tables the number of pairs to be matched should be in the neighborhood of 20. However, the more likely situation to be evaluated is that of an experiment consisting of a series of such matchings of equal numbers of pairs. (In this connection note a caution by the author in *J. Psychol.*, 1940, 53, 614-615, to the effect that the same material should not be matched more than once.) The distribution of the sum of scores obtained is reasonably close to normal when the number of pairs and numbers of matchings are at least as large as 4 and 8, respectively. However, if either number were made much larger, then the other could be made smaller.

† To allow for discontinuity of the data, the deviation is reduced half a unit.

the group of M can be reliably (in the sense of statistically significantly) associated with the correct set of trait descriptions.

Designate the j-th trait of the i-th subject as B_{ij} and the property of the chosen subject (photograph, e.g.) as A. The procedure is as follows: All identifications of traits to subjects are removed and code substituted. The sets of traits are randomized and a judge is asked to preferentially match A with each of the sets $B_{1j}, B_{2j}, \cdots, B_{nj}, j = 1, 2, \cdots, n$. The score s is the sum of numbers assigned to each of the true associates of A in the preferential matching process.

There are two possibilities to consider. The chosen traits have. (a) assignments uncorrelated by pairs, or (b) assignments correlated by pairs. Let ρ_{kl} be the theoretical correlation between assignment score of traits of k and l.

Under the assumption of equi-probability that each of the traits B_{ij} corresponding to A be assigned any value from *1* to M, we may utilize the results of case (i). This gives the mean and variance of a single cell of the matrix of assignments to be $(M + 1)/2$ and $(M^2 - 1)/12$, respectively. Then applying formulas for the mean and variance of a linear expression of variates,

$$\text{mean score} = n(M + 1)/2,$$

$$\mu_{2:s} = \text{variance of score} = \left(n + 2\sum_{i<j}^{1,2,\cdots,n} \rho_{ij}\right)\left(\frac{M^2 - 1}{12}\right)^{*}.$$

If the trait scores may be considered as uncorrelated, then

$$\rho_{ij} = 0 \quad \text{and}$$

$$\mu_{2:s} = n(M^2 - 1)/12.$$

For convenience, a form matrix for scoring is shown in the illustration below in which B_j indicates the j-th trait. Under each B_j is the column of values preferentially assigned to the j-th trait of each of the subjects. The matrix can of course be filled in only after the traits have been decoded.

Illustration: Let $M = 6$, $n = 3$, and suppose that the selected subject was in actuality the second one. Let the completed matrix have the following form:

* It is recognized that estimates, only, of ρ_{ij} may be known, and error thereby introduced into the use of $\mu_{2:s}$.

B_1	B_2	B_3
3	1	5
5	3	6
1	6	1
4	4	2
6	2	4
2	5	3

Assuming $\rho_{ij} = 0$, then the mean $= 10.5$, $\mu_{2:s} = 8.75$, $\sigma_s = 2.96$, $s = 14$. Critical ratio $= (13.5 - 10.5)/2.96 = 1.01$. Referring to the normal tables, the probability that a score as large as 14 would have been obtained equals .16.*

* Under the assumption that $\rho_{ij} = 0$, the same general restrictions as to sizes of n and M for using the normal tables hold as for case (i). I am unable to give any exact statements and proofs of minimum values for n and M for given desirable approximations to true probabilities when using normal tables. However, from examining the values of α_3 and α_4 for selected values of n, M, I would select $n = 3$ and $M = 6$ as minimum values for these constants. Of course increasing n would permit the decreasing of M.

REFERENCES

1. Stuart, C. E. An ESP test with drawings. *J. Parapsychology*, 1942, 6, 20-43.
2. Whitworth, W. A. Choice and chance. 1942 reprint, p. 102, G. E. Stechert & Co., New York.

PSYCHOMETRIKA—VOL. 8, NO. 3
SEPTEMBER, 1943

TABLES OF THE STANDARD ERROR OF TETRACHORIC CORRELATION COEFFICIENT

SAMUEL P. HAYES, JR.

Tables are given for $\sigma_r\sqrt{N}$ for the tetrachoric correlation coefficient for the following values of the correlation in the population: .00, ± .10, ± .20, ⋯ , ± .80, ± .90, ± .95.

The tetrachoric coefficient of correlation is an easily understood and widely applicable measure of relationship. In particular, it is valuable for the measurement of relationship between variables which are not themselves quantitatively measurable. The labor of computing this coefficient has been overcome by the development of certain short-cut methods of approximation which make possible its rapid computation with a degree of precision which is adequate for all ordinary purposes (1, 8).

Wide use of the tetrachoric coefficient has been hindered, however, by the difficulty of computing its standard error. The formula for this standard error was originally developed by Pearson (6).* It is as follows:

$$\sigma_{r_t} = \frac{1}{\chi_0\sqrt{N}}\left\{\frac{(a+d)(b+c)}{4N^2} + \Psi^2_2\frac{(a+c)(b+d)}{N^2} + \Psi^2_1\frac{(a+b)(c+d)}{N^2} + 2\Psi_1\Psi_2\frac{(ad-bc)}{N^2} - \Psi_2\frac{(ab-cd)}{N^2} - \Psi_1\frac{(ac-bd)}{N^2}\right\}^{1/2}$$

where Ψ_1 = area of the normal curve between $\frac{x}{\sigma} = 0$ and $\frac{x}{\sigma} = \beta_1$,

Ψ_2 = area of the normal curve between $\frac{y}{\sigma} = 0$ and $\frac{y}{\sigma} = \beta_2$,

* In Pearson's original article, χ_0 was erroneously printed under the radical instead of outside it. The full formula is given correctly in a later article by Pearson (5). In a recent text by Peters and Van Voorhis (9), the full formula is correctly given, but the meanings of β_1 and β_2 are transposed from their meanings in Pearson's original formula.

$$\beta_1 = \frac{h - rk}{\sqrt{1 - r^2}}, \qquad \beta_2 = \frac{k - rh}{\sqrt{1 - r^2}},$$

(Note that, in using this formula, β_1 or β_2, with the corresponding Ψ, may be *negative*.)

$$\frac{1}{\chi_0} = 2\pi(\sqrt{1 - r^2})\ (e^{\frac{h^2+k^2-2rhk}{2(1-r^2)}}),$$

$h = \frac{x}{\sigma}$ for normal curve divided into $(a + c)$ and $(b + d)$,

$k = \frac{y}{\sigma}$ for normal curve divided into $(a + b)$ and $(c + d)$.

(Note that h and k must always be *positive*, i.e., $(a + c)$ must be arranged to be greater than $(b + d)$ and $(a + b)$ to be greater than $(c + d)$. Also, if r is negative with respect to cells a and d, its sign must be used in the above formula.)

As an alternative to the full formula for the standard error of tetrachoric r, Pearson (5) developed an approximation formula which is much easier to compute. This formula is as follows:

$$\sigma_{r_t} = \frac{\sqrt{(a + b)(c + d)(a + c)(b + d)}}{zz'\sqrt{N}} \left\{(1 - r^2)\left[1 - \left(\frac{\sin^{-1} r}{90^\circ}\right)^2\right]\right\}^{\frac{1}{2}}.$$

In order to simplify computation further, he and other statisticians have computed the values of the function within the bracket for various values of r_t, and have presented them in tables (3, 4, 5, 7).*

Pearson (7) states that the value given by this formula for the standard error "is only approximate and may diverge considerably from the true value for extreme dichotomies." Unfortunately, however, this approximation formula has been adopted for general use without observing the degree of caution displayed by Pearson himself. For example, a recent article (3) states that "the standard error of a tetrachoric r is *given*" (italics mine) by this approximation formula. The statement could have been phrased more properly as "the standard error of a tetrachoric r is approximated by this formula in many cases, especially where the value of r_t does not exceed .80 and where the dichotomy of one variable is fairly similar to the dichotomy of the other."

* In one article (3), the values tabled by Pearson (7) have been recomputed without reference to their earlier publication. When compared with Pearson's tables, the later tables show several discrepancies in the last figure of the tabled values.

The fact is that the standard error computed by the approximation formula may diverge *very greatly* from the standard error computed from the full formula. If r_t is negative, the true value of σ_{r_t} is usually considerably larger than the approximated value, except where one of the variables is divided into fairly equal parts. And if there is a marked difference between the dichotomy of one variable and the dichotomy of the other, the true value of σ_{r_t} for high values of r_t may be much larger than the approximated value even where neither dichotomy is extreme (see Table 1, Part A). The difference between these two values increases as the value of r_t increases. In fact, paradoxically, certain high values of r_t may have *larger* standard errors than lower values, even where the dichotomies of both variables in one case are the same as those in the other. Finally, the true value and the approximated value are quite similar for positive r_t in the case of even extreme dichotomies if the dichotomies of both variables are equally extreme (Table 1, Part B).

TABLE 1

Examples of Values of σ_{r_t} as Determined by the Full Formula and by the Approximation Formula (It is assumed that $N = 100$ in every case.)

Part A Variable *A* split 50-50 Variable *B* split 84-16			*Part B* Variable *A* split 95-5 Variable *B* split 95-5			
r_t	Approximated Value, Positive or Negative r_t	True Value, Positive or Negative r_t	r_t	Approximated Value, Positive or Negative r_t	True Value Positive r_t	True Value Negative r_t
.00	.19	.19	.00	.42	.42	.42
.10	.19	.19	.10	.42	.39	.45
.20	.18	.18	.20	.41	.36	.50
.30	.18	.18	.30	.39	.33	.56
.40	.17	.17	.40	.37	.31	.64
.50	.15	.16	.50	.34	.27	.83
.60	.14	.15	.60	.31	.24	1.26
.70	.12	.13	.70	.26	.20	2.62
.80	.09	.11	.80	.20	.15	13.6
.90	.06	.10	.90	.13	.09	—
.95	.04	.16	.95	.08	.06	—

In order to use the tables presented in this paper, the data of the tetrachoric table being examined should be reduced to percentages and arranged so that the smaller column is on the left and the smaller row

is at the top. Figure 1 is arranged in this way. Note that $a + b < c + d$ and $a + c < b + d$.

10%	20%	30%
10%	60%	70%
20%	80%	100%

a	b	$a + b$
c	d	$c + d$
$a + c$	$b + d$	N

FIGURE 1

Proper Arrangement of Data Preparatory to Computation of Standard Error of Tetrachoric Coefficient

The following tables give values of $\sigma_{r_t}\sqrt{N}$ for certain positive and negative values of r_t for certain dichotomies of each variable, up to the point where the dichotomic lines fall 2.6 σ/x from the means (i.e., where the dividing line leaves only 1/2 of 1% in the tail of the distribution of either or both variables). Rectilinear interpolation within and between these tables will give approximate values which are very close to the true values and which are quite adequate for all ordinary purposes. Graphic curvilinear interpolation will give values even closer to the true values. Of course, in order to translate the value of $\sigma_{r_t}\sqrt{N}$ into σ_{r_t}, the tabled values must be divided by the square root of N for the data used in the particular case.

Separate parts of each table give values of $\sigma_{r_t}\sqrt{N}$ for positive and negative values of r_t. The sign of r_t is determined by the relative volume of that cell of the tetrachoric table that falls in the smaller column and the smaller row of the tetrachoric table. In other words, it is determined by the relative volume of cell a in a table arranged as in Figure 1. The theoretically expected volume of cell a is, of course, $\frac{(a+b)(a+c)}{N}$. If the actual volume of cell a is *greater* than the theoretically expected volume, r_t is positive. If the actual volume of cell a is *less* than the theoretically expected volume, r_t is negative. The value of σ_{r_t} for positive r_t differs from that for negative r_t, except where one or both of the variables is divided into equal parts, in which case the two values are identical.

These tables were computed by using the tables given by Pearson (8) of the volume of one cell of a tetrachoric table. In fairly extreme cases, the values given here may be in error by one or two points in the third significant figure. Furthermore, because the Pearson tables contained no entries for very extreme cases, standard errors were not computed for these cases and the tables given here are therefore incomplete. The cases omitted are so extreme, however, that the utility

of the tables will not be seriously impaired by their omission. If occasion arises to compute them, the tables of tetrachoric functions given in Pearson (7, 8), and certain recently computed tables of other functions (10, 11, 12), will be found very useful.

TABLE 2

$\sigma_{r_t}\sqrt{N}$ for $r_t = .00$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
	y/σ x/σ =	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = \pm .00$					
50%	0.0	1.57	1.64	1.89	2.57	3.46	6.28
30.9%	0.5	1.64	1.72	1.98	2.69	3.62	6.58
15.9%	1.0	1.89	1.98	2.28	3.10	4.17	7.57
5.48%	1.6	2.57	2.69	3.10	4.21	5.67	10.29
2.28%	2.0	3.46	3.62	4.17	5.67	7.63	13.85
0.466%	2.6	6.28	6.58	7.57	10.29	13.85	25.15

TABLE 3

$\sigma_{r_t}\sqrt{N}$ for $r_t = +.10$ and $r_t = -.10$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
	y/σ x/σ =	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = +.10$					
50%	0.0	1.56	1.63	1.88	2.56	3.45	6.30
30.9%	0.5	1.63	1.70	1.95	2.63	3.52	6.35
15.9%	1.0	1.88	1.95	2.22	2.96	3.94	7.02
5.48%	1.6	2.56	2.63	2.96	3.91	5.15	9.04
2.28%	2.0	3.45	3.52	3.94	5.15	6.74	11.70
0.466%	2.6	6.30	6.35	7.02	9.04	11.70	19.82
		$r_t = -.10$					
50%	0.0	1.56	1.63	1.88	2.56	3.45	6.30
30.9%	0.5	1.63	1.72	1.99	2.74	3.72	6.86
15.9%	1.0	1.88	1.99	2.32	3.21	4.42	8.31
5.48%	1.6	2.56	2.74	3.21	4.55	6.30	12.00
2.28%	2.0	3.45	3.72	4.42	6.30	8.80	16.97
0.466%	2.6	6.30	6.86	8.31	12.00	16.97	33.46

TABLE 4

$\sigma_{r_t}\sqrt{N}$ for $r_t = +.20$ and $r_t = -.20$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
y/σ	$x/\sigma =$	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = +.20$					
50%	0.0	1.53	1.60	1.85	2.53	3.43	6.35
30.9%	0.5	1.60	1.66	1.89	2.55	3.41	6.15
15.9%	1.0	1.85	1.89	2.13	2.82	3.72	6.55
5.48%	1.6	2.53	2.55	2.82	3.63	4.71	8.05
2.28%	2.0	3.43	3.41	3.72	4.71	6.02	10.1
0.466%	2.6	6.35	6.15	6.55	8.05	10.1	16.3
		$r_t = -.20$					
50%	0.0	1.53	1.60	1.85	2.53	3.43	6.35
30.9%	0.5	1.60	1.69	1.98	2.77	3.82	7.25
15.9%	1.0	1.85	1.98	2.35	3.37	4.73	9.24
5.48%	1.6	2.53	2.77	3.37	4.98	7.17	14.6
2.28%	2.0	3.43	3.82	4.73	7.17	10.5	21.9
0.466%	2.6	6.35	7.25	9.24	14.6	21.9	48.1

TABLE 5

$\sigma_{r_t}\sqrt{N}$ for $r_t = +.30$ and $r_t = -.30$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
y/σ	$x/\sigma =$	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = +.30$					
50%	0.0	1.47	1.54	1.79	2.48	3.40	6.46
30.9%	0.5	1.54	1.59	1.81	2.44	3.27	5.98
15.9%	1.0	1.79	1.81	2.02	2.65	3.49	6.13
5.48%	1.6	2.48	2.44	2.65	3.35	4.28	7.22
2.28%	2.0	3.40	3.27	3.49	4.28	5.38	8.71
0.466%	2.6	6.46	5.98	6.13	7.22	8.71	13.6
		$r_t = -.30$					
50%	0.0	1.47	1.54	1.79	2.48	3.40	6.46
30.9%	0.5	1.54	1.65	1.95	2.79	3.94	7.82
15.9%	1.0	1.79	1.95	2.37	3.58	5.15	10.7
5.48%	1.6	2.48	2.79	3.58	5.58	8.46	18.8
2.28%	2.0	3.40	3.94	5.15	8.46	13.1	30.8
0.466%	2.6	6.46	7.82	10.7	18.8	30.8	76.4

TABLE 6

$\sigma_{r_t}\sqrt{N}$ for $r_t = +.40$ and $r_t = -.40$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
	y/σ x/σ =	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = +.40$					
50%	0.0	1.39	1.46	1.70	2.40	3.37	6.71
30.9%	0.5	1.46	1.50	1.70	2.31	3.12	5.84
15.9%	1.0	1.70	1.70	1.89	2.46	3.24	5.73
5.48%	1.6	2.40	2.31	2.46	3.05	3.87	6.50
2.28%	2.0	3.37	3.12	3.24	3.87	4.78	7.64
0.466%	2.6	6.71	5.84	5.73	6.50	7.64	11.4
		$r_t = -.40$					
50%	0.0	1.39	1.46	1.70	2.40	3.37	6.71
30.9%	0.5	1.46	1.57	1.90	2.82	4.12	8.80
15.9%	1.0	1.70	1.90	2.39	3.78	5.80	13.3
5.48%	1.6	2.40	2.82	3.78	6.45	10.6	26.9
2.28%	2.0	3.37	4.12	5.80	10.6	18.2	50.5
0.466%	2.6	6.71	8.80	13.3	26.9	50.5	151.

TABLE 7

$\sigma_{r_t}\sqrt{N}$ for $r_t = +.50$ and $r_t = -.50$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
	y/σ x/σ =	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = +.50$					
50%	0.0	1.28	1.35	1.60	2.32	3.37	7.24
30.9%	0.5	1.35	1.38	1.57	2.15	2.96	5.79
15.9%	1.0	1.60	1.57	1.72	2.24	2.95	5.34
5.48%	1.6	2.32	2.15	2.24	2.73	3.44	5.73
2.28%	2.0	3.37	2.96	2.95	3.44	4.19	6.60
0.466%	2.6	7.24	5.79	5.34	5.73	6.60	9.56
		$r_t = -.50$					
50%	0.0	1.28	1.35	1.60	2.32	3.37	7.24
30.9%	0.5	1.35	1.48	1.83	2.89	4.47	10.7
15.9%	1.0	1.60	1.83	2.42	4.21	7.00	18.6
5.48%	1.6	2.32	2.89	4.21	8.32	15.2	45.6
2.28%	2.0	3.37	4.47	7.00	15.2	29.0	95.0
0.466%	2.6	7.24	10.7	18.6	45.6	95.0	—

TABLE 8

$\sigma_{r_t}\sqrt{N}$ for $r_t = +.60$ and $r_t = -.60$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
y/σ	$x/\sigma =$	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = +.60$					
50%	0.0	1.15	1.21	1.46	2.24	3.46	8.50
30.9%	0.5	1.21	1.23	1.40	1.96	2.79	5.94
15.9%	1.0	1.46	1.40	1.53	1.98	2.64	4.99
5.48%	1.6	2.24	1.96	1.98	2.38	2.98	5.00
2.28%	2.0	3.46	2.79	2.64	2.98	3.59	5.61
0.466%	2.6	8.50	5.94	4.99	5.00	5.61	7.88
		$r_t = -.60$					
50%	0.0	1.15	1.21	1.46	2.24	3.46	8.50
30.9%	0.5	1.21	1.36	1.76	3.04	5.21	15.*
15.9%	1.0	1.46	1.76	2.52	5.09	9.64	32.2
5.48%	1.6	2.24	3.04	5.09	12.6	26.8	110.
2.28%	2.0	3.46	5.21	9.64	26.8	61.2	214.
0.466%	2.6	8.50	15.*	32.2	110.	214.	—

* By curvilinear interpolation.

TABLE 9

$\sigma_{r_t}\sqrt{N}$ for $r_t = +.70$ and $r_t = -.70$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
y/σ	$x/\sigma =$	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = +.70$					
50%	0.0	.97	1.04	1.30	2.21	3.85	12.1
30.9%	0.5	1.04	1.05	1.20	1.75	2.65	6.68
15.9%	1.0	1.30	1.20	1.29	1.68	2.30	4.77
5.48%	1.6	2.21	1.75	1.68	1.98	2.47	4.27
2.28%	2.0	3.85	2.65	2.30	2.47	2.94	4.60
0.466%	2.6	12.1	6.68	4.77	4.27	4.60	6.27
		$r_t = -.70$					
50%	0.0	.97	1.04	1.30	2.21	3.85	12.1
30.9%	0.5	1.04	1.21	1.70	3.57	7.23	—
15.9%	1.0	1.30	1.70	2.85	7.52	17.8	—
5.48%	1.6	2.21	3.57	7.52	26.2	71.5	806.
2.28%	2.0	3.85	7.23	17.8	71.5	—	—
0.466%	2.6	12.1	—	—	806.	—	—

TABLE 10

$\sigma_{r_t}\sqrt{N}$ for $r_t = +.80$ and $r_t = -.80$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
	y/σ x/σ =	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = +.80$					
50%	0.0	.76	.83	1.12	2.44	5.57	30.4
30.9%	0.5	.83	.81	.95	1.53	2.73	10.0
15.9%	1.0	1.12	.95	.99	1.33	1.94	5.07
5.48%	1.6	2.44	1.53	1.33	1.51	1.90	3.63
2.28%	2.0	5.57	2.73	1.94	1.90	2.21	3.53
0.466%	2.6	30.4	10.0	5.07	3.63	3.53	4.75
		$r_t = -.80$					
50%	0.0	.76	.83	1.12	2.44	5.57	30.4
30.9%	0.5	.83	1.04	1.79	5.66	19.*	—
15.9%	1.0	1.12	1.79	4.20	19.4	75.4	—
5.48%	1.6	2.44	5.66	19.4	136.	—	—
2.28%	2.0	5.57	19.*	75.4	—	—	—
0.466%	2.6	30.4	—	—	—	—	—

* By curvilinear interpolation.

TABLE 11

$\sigma_{r_t}\sqrt{N}$ for $r_t = +.90$ and $r_t = -.90$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
	y/σ x/σ =	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = +.90$					
50%	0.0	.48	.55	1.02	5.46	30.5	—
30.9%	0.5	.55	.51	.63	1.55	4.89	—
15.9%	1.0	1.02	.63	.62	.91	1.74	7.46
5.48%	1.6	5.46	1.55	.91	.93	1.21	3.11
2.28%	2.0	30.5	4.89	1.74	1.21	1.35	2.35
0.466%	2.6	—	—	7.46	3.11	2.35	2.75
		$r_t = -.90$					
50%	0.0	.48	.55	1.02	5.46	30.5	—
30.9%	0.5	.55	.96	3.20	14.4	—	—
15.9%	1.0	1.02	3.20	6.97	—	—	—
5.48%	1.6	5.46	14.4	—	—	—	—
2.28%	2.0	30.5	—	—	—	—	—
0.466%	2.6	—	—	—	—	—	—

TABLE 12

$\sigma_{r_t}\sqrt{N}$ for $r_t = +.95$ and $r_t = -.95$

% of combined groups falling in smaller group:		% of combined groups above criterion:					
		50%	30.9%	15.9%	5.48%	2.28%	0.466%
y/σ	$x/\sigma =$	0.0	0.5	1.0	1.6	2.0	2.6
		$r_t = +.95$					
50%	0.0	.30	.40	1.62	—	—	—
30.9%	0.5	.40	.32	.45	2.94	—	—
15.9%	1.0	1.62	.45	.38	.71	2.63	—
5.48%	1.6	—	2.94	.71	.57	.80	4.94
2.28%	2.0	—	—	2.63	.80	.82	1.78
0.466%	2.6	—	—	—	4.94	1.78	1.66
		$r_t = -.95$					
50%	0.0	.30	.40	1.62	—	—	—
30.9%	0.5	.40	1.52	30.6	—	—	—
15.9%	1.0	1.62	30.6	—	—	—	—
5.48%	1.6	—	—	—	—	—	—
2.28%	2.0	—	—	—	—	—	—
0.466%	2.6	—	—	—	—	—	—

REFERENCES

1. Cheshire, L., Saffir, M., and Thurstone, L. L. Computing diagrams for the tetrachoric correlation coefficient. Chicago: Univ. Chicago Bookstore, 1933.
2. Davenport, C. B. and Ekas, M. P. Statistical methods in biology, medicine, and psychology. New York: John Wiley & Sons, 1936.
3. Guilford, J. P., and Lyons, Thoburn C. On determining the reliability and significance of a tetrachoric coefficient of correlation. *Psychometrika*, 1942, 7, 243-249.
4. Kelley, Truman L. Statistical methods. New York: Macmillan, 1924, p. 258.
5. Pearson, Karl. On the probable error of a coefficient of correlation as found from a fourfold table. *Biometrika*, 1913, 9, 22-28.
6. Pearson, Karl. On the correlation of characters not quantitatively measurable. *Philos. Trans.*, Series A, 1900, 195, 1-47.
7. Pearson, Karl (editor). Tables for statisticians and biometricians. Part I, Cambridge: Cambridge University Press, 1914, pp. xl-xli and p. 35.
8. Pearson, Karl (editor). Tables for statisticians and biometricians. Part II, Cambridge: Cambridge University Press, 1931.
9. Peters, Charles, and Van Voorhis, Walter R. Statistical procedures and their mathematical bases. New York: McGraw-Hill, 1940, 366-384.
10. Works Progress Administration, Mathematical Tables Project. Tables of the exponential function. Washington: National Bureau of Standards, 1942.
11. Works Progress Administration, Mathematical Tables Project. Tables of natural logarithms. Volumes I and II, Washington: National Bureau of Standards, 1941.
12. Works Progress Administration, Mathematical Tables Project. Tables of probability functions. Volume II, Washington: National Bureau of Standards, 1942.